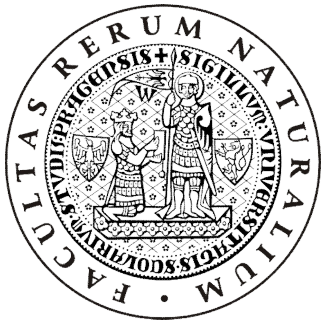
Charles University in Prague

Faculty of Science

Study program: Biology

Study field: Cell and developmental biology



Bc. Viola Hausnerová

Spliceosome assembly

Formování sestřihového komplexu

Diploma thesis

Supervisor: Mgr. David Staněk, PhD.

Prague, 2011

Declaration

I declare that I compiled the thesis on my own and that I properly cited all sources of information and literature. Neither this work nor any part of it was submitted to gain any different or equal academic title.

Prague 29. 8. 2011

Viola Hausnerová

Acknowledgement

I would like to thank my supervisor David Staněk for the time he spent with me and for his advice and comments on this work and on all the work I did during my stay in his lab. I would also like to thank all the members of RNA biology department of Institute of Molecular genetics AS CR for creating a friendly atmosphere, helping me with my experiments and being supportive all the time.

I would especially like to thank Ivan Novotný and Jarmila Hnilicová for their great help with performance of FRET and RIP and plenty of advice they gave to me. I would never be able to complete this work without their contribution.

Table of contents

Abstract **1**
Abstrakt **2**
List of abbreviations **3**
1. Introduction **5**
2. Aims of the diploma thesis **6**
3. Review of literature **7**
3.1 Spliceosome composition and structure 7
3.1.1 U1 snRNP 8
3.1.2 U2 snRNP 9
3.1.3 U5 snRNP 9
3.1.4 U4/U6 snRNP (di-snRNP) 11
3.2 snRNPs biogenesis 12
3.2.1 RNA polymerase II transcribed snRNAs 12
3.2.2 RNA polymerase III transcribed snRNAs 14
3.3 Spliceosome assembly 16
3.3.1 Splice sites and BPS recognition (E complex formation) 16
3.3.2 A complex formation 18
3.3.3 Splice sites cooperation within E and A complex 19
3.3.4 Tri-snRNP association (B and C complex formation) 25
3.3.5 Penta-snRNP versus stepwise assembly 28
4. Materials and methods **30**
4.1 Materials 30
4.1.1 List of instruments 30
4.1.2 Cell lines 30
4.1.3 Bacterial strains 31
4.1.4 Antibodies 31
4.1.5 Plasmids 31
4.1.6 Primers 33
4.2 Methods 34
4.2.1 Restriction digestions and ligation 34
4.2.2 Transformation of competent bacteria 35
4.2.3 Plasmid DNA preparation from bacterial culture 36
4.2.4 Transfection of HeLa cells 36
4.2.5 Site directed mutagenesis 36
4.2.6 Total RNA isolation 37
4.2.7 DNAse treatment 38
4.2.8 Reverse transcription 38
4.2.9 RT-PCR 39
4.2.10 Quantitative PCR (qPCR) 39
4.2.11 Cell lysate preparation (for RNA immunoprecipitation) 40
4.2.12 Immunoprecipitation (for RNA immunoprecipitation) 40
4.2.13 RNA isolation from sepharose beads 41
4.2.14 SDS PAGE 41
4.2.15 Western blotting 42
4.2.16 Horizontal agarose gel electrophoresis 43
4.2.17 Urea RNA gel 43

4.2.19 Fluorescence resonance energy transfer (FRET) 44
5. Results 46
5.1 Splicing of β-globin pre-mRNA 46
5.2 The E3 transcripts are bound by MS2 protein with high affinity 53
5.3 Differential splice sites occupation detected by FRET 57
5.4 RIP analysis of splice sites attached complexes 59
6. Discussion 70
7. References 76

Abstract

Pre-mRNA splicing is a process in which introns are removed from eukaryotic transcripts and exons are ligated together. Splicing is catalyzed by spliceosome, a large ribonucleoprotein complex composed of five small nuclear RNAs and more than 100 additional proteins, which recognizes 5′ splice site, branch point site and 3′ splice site and performs two transesterification reactions to produce mRNA molecules. 5′ splice site is recognized by U1 snRNP and U2 auxiliary factor (U2AF) is involved in branch point and 3′ splice site recognition in the early splicing complex. There is some evidence of splice sites cooperation during intron recognition *in vitro* but little is known about the situation *in vivo*. Using Fluorescence resonance energy transfer (FRET) and RNA immunoprecipitation (RIP) methods, we have investigated the early stages of spliceosome assembly. We have employed splicing reporters based on β-globin gene and MS2 stem loops to detect interactions of proteins on RNA molecule directly in the cell nucleus. Results of FRET indicate that intact 5′ splice site is required for U2AF35 interaction with 3′ splice site and that U1C recruitment to 5′ splice site is partially limited upon 3′ splice site mutation. We have also confirmed by RIP that U2 snRNP association with pre-mRNA molecule requires presence of 5′ splice site. Our results thus bring the first evidence of *in vivo* cooperation of spliceosomal components in the process of intron recognition.

Key words: pre-mRNA splicing, spliceosome assembly, FRET, U1 snRNP, U2AF, U2 snRNP

Abstrakt

Introny jsou vystřiženy z eukaryotických transkriptů a exony jsou spojeny dohromady během procesu zvaného sestřih pre-mRNA. Sestřih je katalyzován sestřihovým komplexem. Jedná se o velký ribonukleoproteinový komplex složený z pěti malých jaderných RNA a více než stovky proteinů. Tento komplex rozpoznává 5´ sestřihové místo, místo větvení a 3´ sestřihové místo a následně provádí dvě transesterifikační reakce, jejichž výsledkem je zralá molekula mRNA. V rámci časného sestřihového komplexu je 5´ sestřihové místo definováno pomocí U1 snRNP a na rozpoznání místa větvení a 3´ sestřihového místa se podílí U2 pomocný faktor (U2 auxiliary factor, U2AF). Spolupráce sestřihových míst byla částečně popsána *in vitro*, ale situace *in vivo* není dosud zcela objasněna. V této studii jsme použili fluorescenční rezonanční energetický transfer (Fluorescence resonance energy transfer, FRET) a RNA imunoprecipitaci (RIP) k popsání časných kroků skládání sestřihového komplexu. Abychom detekovali interakce proteinů na RNA molekule přímo v buněčném jádře, uplatnili jsme sestřihové reportéry kódující β-globinový gen a vlásenky z fága MS2. Výsledky FRETu ukazují, že intaktní 5´ sestřihové místo je vyžadováno pro vazbu U2AF35 na 3´ sestřihové místo a že vazba U1C je částečně omezena v přítomnosti mutace 3´ sestřihového místa. Dále jsme pomocí RIPu prokázali, že vazba U2 snRNP na pre-mRNA je podmíněna přítomností 5´ sestřihového místa. Naše výsledky jsou prvním *in vivo* důkazem spolupráce komponent sestřihového komplexu během rozpoznání intronu.

Klíčová slova: sestřih pre-mRNA, formování sestřihového komplexu, FRET, U1 snRNP, U2AF, U2 snRNP

List of abbreviations

3ss	3´ splice site
5ss	5´ splice site
APS	ammonium persulfate
ATP	adenosine triphosphate
BBP	branch point binding protein
Bp	base pairs
BPS	branch point site
BSA	bovine serum albumin
CBP	Cap binding complex
CFP	Cyan fluorescent protein
ChIP	Chromatin immunoprecipitation
CMV	Cytomegalovirus
CPSF	Cleavage and polyadenylation specificity factor
DABCO	1,4-diazabicyclo[2.2.2]octane
DMEM	Dulbecco´s modified Eagle medium
EJC	Exon junction complex
ESE	exonic splicing enhancer
ESS	exonic splicing silencer
FBP	Formin binding protein
FCS	Fluorescence correlation spectroscopy
FRAP	Fluorescence recovery after photobleaching
FRET	Fluorescence resonance energy transfer
GFP	Green fluorescent protein
hnRNP	heterogeneous nuclear ribonucleoprotein
IP	immunoprecipitation
ISE	intronic splicing enhancer
ISS	intronic splicing silencer
nt	nucleotides
NTC	Nineteen complex
PBS	phosphate buffer saline
PBST	phosphate buffer saline - Tween

PCR	Polymerase chain reaction
PIPES	piperazine-N,N′-bis(2-ethanesulfonic acid)
PolII	RNA polymerase II
PolIII	RNA polymerase III
PPT	polypyrimidine tract
pre-mRNA	precursor mRNA
PTB	Polypyrimidine tract binding protein
qPCR	quantitative PCR
RFP	Red fluorescent protein
RIP	RNA immunoprecipitation
RNP	ribonucleoprotein
RRM	RNA recognition motif
rRNA	ribosomal RNA
RS domain	arginine serine rich domain
RT	reverse transcription
RT-PCR	reverse transcription PCR
scaRNA	small Cajal body RNA
SDS	sodium dodecyl sulfate
SDS PAGE	SDS polyacrylamid gel electrophoresis
SF1	Splicing factor 1
SL	stem loop
SMN	Survival of motor neuron protein
snRNA	small nuclear RNA
snRNP	small nuclear ribonucleoprotein
SS	splice site(s)
SRp	SR protein
TBE	tris-borate-EDTA
TEMED	tetramethylethylenediamine
U2AF	U2 auxiliary factor
WB	Western blot
WT	wild type
YFP	Yellow fluorescent protein

1. Introduction

Non-coding sequences which interrupt the coding parts of eukaryotic transcripts need to be spliced out before the mature mRNA is produced. This is mediated by process called pre-mRNA splicing. Taking place in the cell nucleus, it is catalyzed by multifactorial spliceosome complex, one of the largest and the most complicated molecular machines in the cell. Its role is to recognize coding and non-coding sequence regions in the primary transcript and remove the non-coding ones, called introns, and join together the coding ones, called exons. Together with capping and polyadenylation, pre-mRNA splicing belongs to the group of posttranscriptional processing steps which have a distinct role in gene expression regulation.

The spliceosome complex is formed by 5 small nuclear RNAs, so called snRNAs (U1, U2, U4, U5 and U6) which are the main actors of the spliceosomal catalysis. These RNAs are accompanied by more than 100 spliceosomal proteins and variety of non-spliceosomal factors, acting together on introns removal from eukaryotic pre-mRNAs. The small RNAs and proteins are assembled in ribonucleoprotein particles, so called snRNP. These complexes undergo a set of coordinated conformational changes during each cycle of splicing. They are maturated in both nucleus and cytoplasm. During splicing reaction, snRNPs assemble on pre-mRNA template. There are a few possible pathways of spliceosome assembly and up to date, two canonical models were described: stepwise assembly model and penta-snRNP (supraspliceosome) model.

Spliceosome complex is a molecular machinery of high fidelity, able to recognize both constitutive and alternative exons. Nevertheless, our knowledge of splice sites selection and correct intron definition has been restricted to *in vitro* data, based on cell free systems. In this study, the question of *in vivo* assembly pathways was addressed and appropriate approaches were chosen to answer this question. The main goal was to describe spliceosome assembly *in situ*. Using HeLa cell line and set of vectors expressing β-globin pre-mRNA, splicing complex formation was studied by fluorescence microscopy (Fluorescence resonance energy transfer, FRET) and RNA immunoprecipitation (RIP). To decode the impact of splice sites sequences, a system of splice site mutations was established. Using these approaches, a relationship between 5´ and 3´ splice sites in spliceosome assembly was investigated.

2. Aims of the diploma thesis

- to create a set of splicing reporters based on the β-globin gene with mutated splice sites next to MS2 binding sites that will be used for *in situ* detection of early splicing events
- to create cell lines stably expressing the vectors
- to describe RNA-protein interactions during intron recognition directly in the cell nucleus using *in vivo* microscopy techniques

3. Review of literature

3.1 Spliceosome composition and structure

Spliceosome is a multimegadalton ribonucleoprotein (RNP) complex which catalyses intron removal from eukaryotic precursor mRNA (pre-mRNA) transcripts. It consists of nearly 150 proteins (Zhou et al., 2002). In addition to proteins, there are 5 small nuclear RNAs (snRNAs) involved – U1, U2, U4, U5 and U6. Similarly to other RNPs, both RNA and protein elements are joined together in a complex structure undergoing a series of conformational changes and all types of interactions (protein-protein, RNA-protein, RNA-RNA) are present in the structure. It is important to notice that the RNA fraction is likely responsible for the catalytic process of splicing ant the proteins only act as a scaffold for RNA molecules (Wahl et al., 2009). Spliceosome is a highly dynamic entity which exhibits flexible and versatile compositional changes during each cycle of catalysis. It needs to work on variable pre-mRNA substrates which do not share any secondary or tertiary structure. In case of mammalian transcripts, the sequence elements recognized by spliceosome are poorly conserved. In spite of all of these circumstances, spliceosome performs the splicing reaction with high fidelity and in dynamic manner, in coordination with transcription and subsequent mRNA transport and quality control.

Spliceosome is composed in five small ribonucleoprotein particles – snRNPs. This type of composition allows it to act in a dynamic and ordered sequence of both ATP dependent and independent steps by which it recognizes the pre-mRNA substrate, performs the two transesterification reactions and releases the mature mRNA product. During each catalytic cycle, snRNPs become assembled and dissembled from the substrate, which leads to conformational changes of both RNAs and proteins and differential interactions among spliceosomal RNAs itself and among substrate RNA (Wahl et al., 2009).

In this study, the major type of spliceosome catalyzing excision of the most abundant class of introns (U2 type) will be addressed and thus discussed in the review of literature. Neither the principles of minor spliceosome (which catalyses U12 type introns splicing), nor splicing of autocatalytic introns, will be discussed here.

3.1.1 U1 snRNP

U1 snRNP plays a crucial role in 5′ splice site recognition via U1 snRNA and U1C protein which both sequence specifically and without base pairing involvement select the correct 5′ end of an intron. Of the two interactions, the protein one appears to be the earlier one (Du and Rosbash, 2002). The structure was revealed by crystallization of the particle. Human U1 snRNA is formed by 4 helixes, three on the 5′ end and fourth close to the 3′ end. Between them, a ring of seven Sm proteins (Sm-E, Sm-G, Sm-D3, Sm-B, Sm-D1, Sm-D2, Sm-F) is assembled. The ring has a hole in the middle through which the RNA is laid (Stark et al., 2001). The U1C protein contains several α helixes: one of them contacts the Sm ring and another one, which comprises a Zinc finger motif, has a strategic function in stabilizing the base pairing region of U1 snRNA and 5′ splice site of pre-mRNA (Pomeranz Krummel et al., 2009).

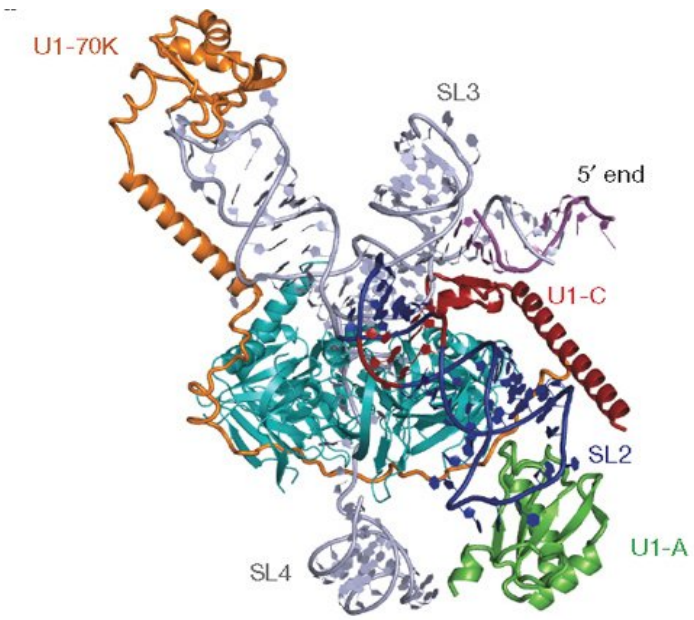


Figure 3.1: RNA and proteins composition within U1 snRNP. The four stem loops (SL1 – SL4), a ring of 7 Sm proteins as well as U1C, U1A and U1-70K form the particle (Pomeranz Krummel et al., 2009).

U1-70K protein, together with components of Sm ring, mediates U1C binding to the particle. There is an RNA binding motif in this protein which allows binding of one of the stem loops of U1 snRNA. The last U1 snRNA component, U1A, is characterized by two RNA binding motifs. One of them is in contact with U1 snRNA stem loop 2. The second one is of unknown function with significant similarities to U2B´´ RNA binding motif, thus might have a role in overall particle stability (Pomeranz Krummel et al., 2009).

3.1.2 U2 snRNP

U2 snRNP functionally contributes to 3´ splice site recognition via its interaction with branch point site, although the 3´ splice site sequence itself is recognized by U2AF [U2 auxiliary factor, which is formed by 35 KDa and 65 KDa subunits; (Wu et al., 1999)]. Same as U1, U4 and U5 snRNA, U2 snRNA is bound by a ring of seven Sm proteins. In terms of the entire RNP particle, two complexes which differ in size were characterized: 12S complex and 17S complex. The smaller one only contains the RNA molecule, Sm ring and two U2 specific proteins: U2A´ and U2B´´, that are both assembled on the 3´ terminal stem loop of the U2 snRNA molecule (Price et al., 1998). The 17S complex is enriched in additional associated splicing factors called SF3a and SF3b, while the first mentioned protein was found to be associated with the 3´ half of the U2 snRNA and the second one with the 5´ half. Both of these proteins comprise several subunits of variable sizes. SF3a and b contribute to overall branch point recognition and interaction with the adenosine nucleotide and have a role in subsequent transition into active spliceosomal complex (Kramer et al., 1999; Spadaccini et al., 2006).

3.1.3 U5 snRNP

U5 snRNP functionally belongs to a bigger particle called tri-snRNP which also contains U4/U6 di-snRNP. U5 snRNP is the largest of all spliceosomal particles. Within tri-snRNP, U5 carries out majority of catalytic activity of spliceosome, forming the catalytic core of the activated molecular complex (Luhrmann and Stark, 2009).

It is important to notice that U5 snRNP exists during a part of the spliceosomal cycle as tri-snRNP component. The protein composition is thus partially altered in separated particle in comparison to the associated form. U5 snRNP is formed by snRNA molecule and a characteristic ring of seven Sm proteins. In addition, a set of U5 specific proteins was identified: Prp8, Brr2, Snu114, Prp6 and Prp28 (Sander et al., 2006). The activity of Prp28 involves ATP-dependent U1 snRNA unwinding from the 5´ splice site (Staley and Guthrie, 1999). Brr2 which possess RNA helicase activity works together with Snu114 GTPase on U4-U6 snRNAs double helix separation. Brr2 is thus responsible for U6 snRNA release from its original structure (Bartels et al., 2002). Prp8 is a large and highly conserved protein with a range of functions in spliceosomal rearrangements involving important sequence elements of pre-mRNA, suggested to carry out the most important catalytic activity of all spliceosomal proteins (Kuhn et al., 2002). It contains an RNA recognition motif and nuclear localization signal. Mutations of Prp8 are known to cause genetic disorders (Grainger and Beggs, 2005).

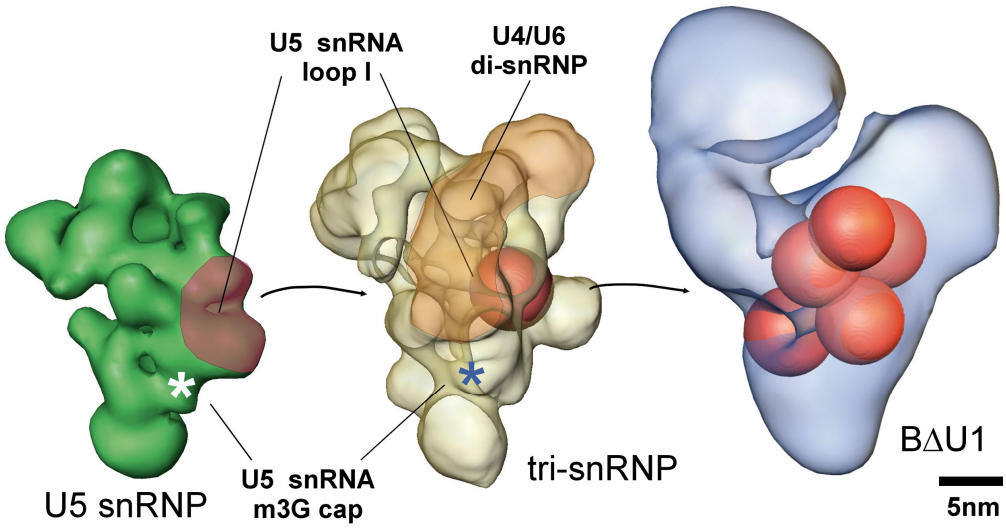


Figure 3.2: The tri-snRNP organization. A model of fitting of U5 snRNP (left) into tri-snRNP particle (middle) and several possible fits (red spheres) into prespliceosomal B complex lacking U1 snRNP (right) (Sander et al., 2006).

The loop 1 of U5 snRNA appears to be localized close to the catalytic core of the spliceosome and is known to play a role in base-pairing with exons to make their ends accessible for the transesterification reaction. U5 snRNA together with di-snRNP forms the central part of tri-snRNP and is likely to contact U2 snRNA during splicing catalysis (Sander et al., 2006).

3.1.4 U4/U6 snRNP (di-snRNP)

In the di-snRNP, the catalytic activity is mediated by U6 snRNP part. The proposed role of U4 snRNP is protecting reactive U6 snRNA from catalytic contacts before the di-snRNP particle is assembled into active spliceosome. Indeed, the U4–U6 base-pairing contact is unwound right after tri-snRNP joins spliceosome. U6 snRNA is then involved in U2 snRNA base-pairing and catalytic contact to 5´splice site region (Will and Luhrmann, 2011).

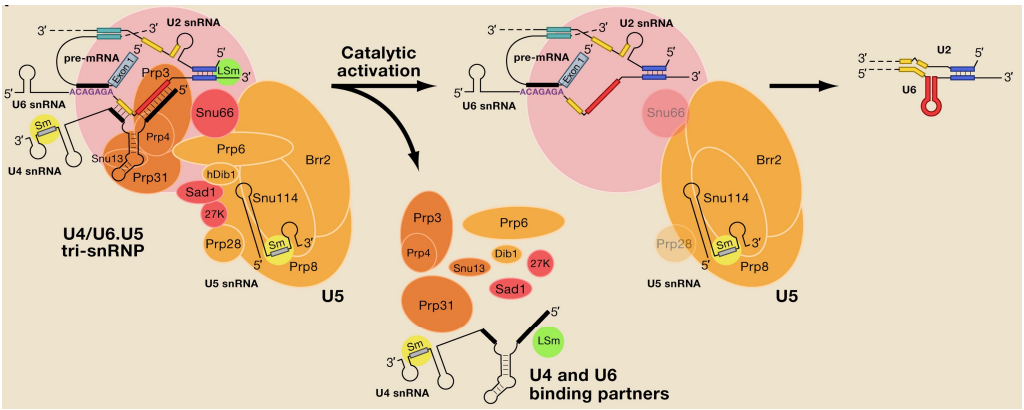


Figure 3.3: The complex of transitions within tri-snRNP particle. Involving three RNA molecules and a large set of proteins, tri-snRNP undergoes several transitions and compositional changes before entering the mature catalytic spliceosome. The two areas of base-pairing between U4 and U6 snRNAs are interrupted to create an interface for U6/U2 snRNAs cooperation during splicing reaction (Wahl et al., 2009).

Regarding the molecular architecture of U4/U6 snRNP, it consists of two snRNA molecules, of which U6 is the exceptional one: in contrast to other snRNAs (RNA polymerase II transcripts), U6 snRNA is transcribed by RNA polymerase III. In addition, it is bound by a ring of Lsm (like Sm) proteins, evolutionary paralogs of Sm proteins (Veretnik et al., 2009). U4 snRNA presents a conventional snRNP carrying a ring of Sm proteins. There are several di-snRNP specific proteins, namely 20K, 60K (Prp4p), 90K (Prp3p), 15.5K (Snu13p) and 61K (Prp31p)[*]. 15.5K protein is required for 20K/60K/90K complex and 61K protein to bind U4/U6 snRNAs. 60K and 90K were suggested to bind U4/U6 snRNAs (Nottrott et al., 2002). U6 snRNA additionally contains a 150 KDa protein called p110 (also known as Sart3; Prp24p in yeast). The main role of this factor is to favor conformation of U6 snRNP which promotes di-snRNP assembly (Karaduman et al., 2008).

3.2 snRNPs biogenesis

snRNPs (small nuclear ribonucleoproteins) are particles which form spliceosome and perform the catalysis of splicing reaction. The RNA and protein composition of these spliceosomal building blocks was described earlier. At this stage, the dynamic process of *in vivo* snRNPs maturation will be discussed.

The small nuclear RNAs (snRNAs) can be divided into two groups: RNA polymerase II (PolII) transcribed – U1, U2, U4 and U5, and RNA polymerase III (PolIII) transcribed – U6. U refers to uridine rich composition of these RNAs. After transcription, snRNAs undergo several modifications before they become a stable part of mature snRNPs.

3.2.1 RNA polymerase II transcribed snRNAs

PolII transcribed snRNA genes are organized in clusters of multiple copies. Similarly to all other PolII transcripts, snRNAs 5´ end is protected by N7-methyl guanosine cap (Eliceiri and Sayavedra, 1976). The 3´ end processing is mediated by Integrator complex which is associated with C terminal domain of PolII

[*] Yeast orthologs are stated in the brackets.

and its composition exhibits some similarities to cleavage and polyadenylation specificity factor (CPSF), which is involved in mRNAs 3′ end processing (Baillat et al., 2005). The processed transcript proceeds to the cytoplasm for further maturation steps. snRNA is transported through the nuclear pore in complex with Cap binding protein (CBP), phosphorylated adaptor for RNA export (PHAX), CRM1 nuclear export receptor and Ran GTPase (Segref et al., 2001).

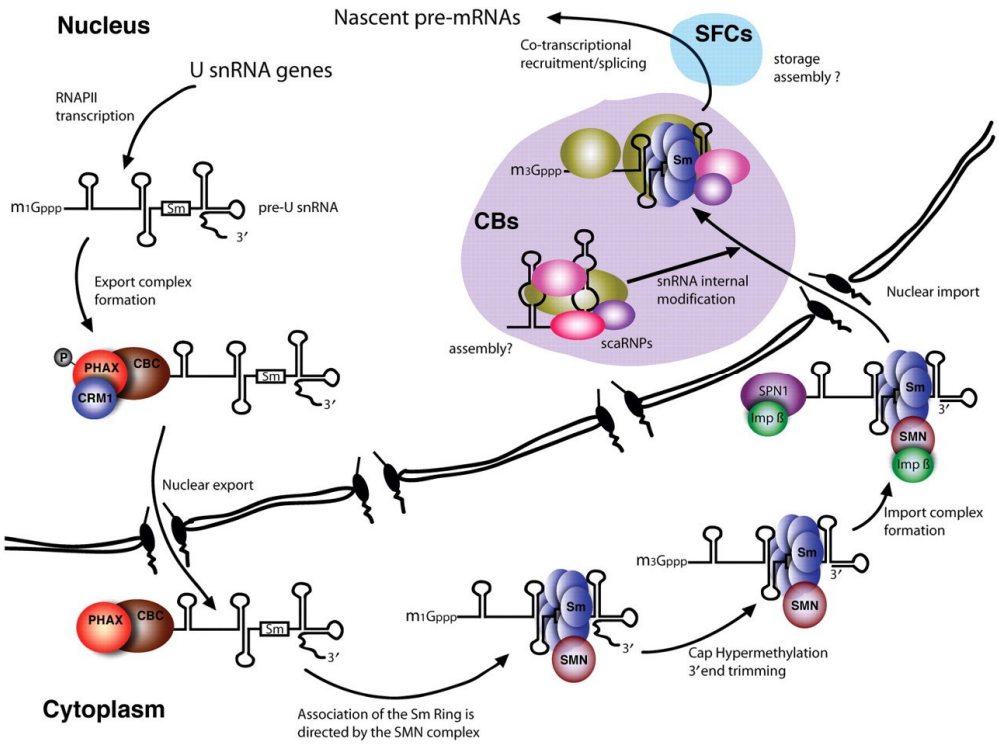


Figure 3.4: The life cycle of PolII transcribed snRNAs. U1, U2, U4 and U5 snRNPs occur in both cytoplasm and nucleus during their maturation process. A set of proteins (PHAX, CBP, CRM1, SMN, snurportin1) as well as RNAs (scaRNAs) are involved it the series of steps leading to mature snRNPs production. SnRNPs move through CBs and eventually also through SFC (splicing factor compartments) within the nucleus (Patel and Bellini, 2008).

During cytoplasmic phase of snRNPs maturation, the ring of Sm proteins is assembled on a conserved region of snRNA. The key factor responsible for the proper ring formation is survival of motor neuron protein (SMN) complex, which is composed of the SMN proteins and seven Gemin proteins (Battle et al., 2006). The hypermethylation of 5′ cap is carried out by SMN associated hypermethylase Tgs1. Subsequent nuclear import is supported by SMN contact with import adaptor Snurportin1. The Sm proteins binding the snRNA and the existence of trimethyl-G cap are sufficient nuclear import signals (Narayanan et al., 2004; Rollenhagen and Pante, 2006).

Modified snRNPs (now containing also the Sm ring, thus snRNPs) coming back to the nucleus are still not completely maturated. Instead, they are exposed to additional series of steps which take place in several subnuclear domains. In the first step, snRNPs localize to Cajal bodies (CBs), non-membrane nuclear suborganelles which appear in certain cell types and cell lines and were reported to play a role in snRNPs biogenesis, maturation and several other nuclear processes, including tri-snRNP recycling during splicing (Stanek et al., 2008). 2′-O-methylation and pseudouridylation of snRNAs are carried out in CBs with assistance of guide scaRNAs (small Cajal body RNAs) which, in a sequence dependent manner, determine the nucleotides of snRNAs to be modified by adequate enzymes. Thus, targeting of snRNPs to CBs is a necessary step in the maturation process to produce fully functional particles (Darzacq et al., 2002; Kiss et al., 2002). The assembly of snRNPs specific proteins is believed to be processed in the nucleus [reviewed in (Stanek and Neugebauer, 2006)].

3.2.2 RNA polymerase III transcribed snRNAs

This group only consists of U6 snRNA which is in many ways an exception among snRNAs. In terms of snRNP maturation, there is not any cytoplasmic phase. After PolIII transcript generation, transcription is terminated by polyU sequence which is a typical feature of PolIII transcripts. 5′ and 3′ ends of the nascent transcript are then bound by La proteins which target U6 to the nucleolus (Wolin and Cedervall, 2002). La proteins are then exchanged for a ring of seven Lsm proteins (2 – 8). The modifications of certain nucleotides, same as above mentioned

modifications of PolII transcribed snRNAs, were localized to the nucleolus (Lange and Gerbi, 2000). In contrast to PolII transcribed snRNAs, this set of modifications is directed by snoRNAs (small nucleolar RNAs), the ones which participate on rRNAs modifications (Ganot et al., 1999).

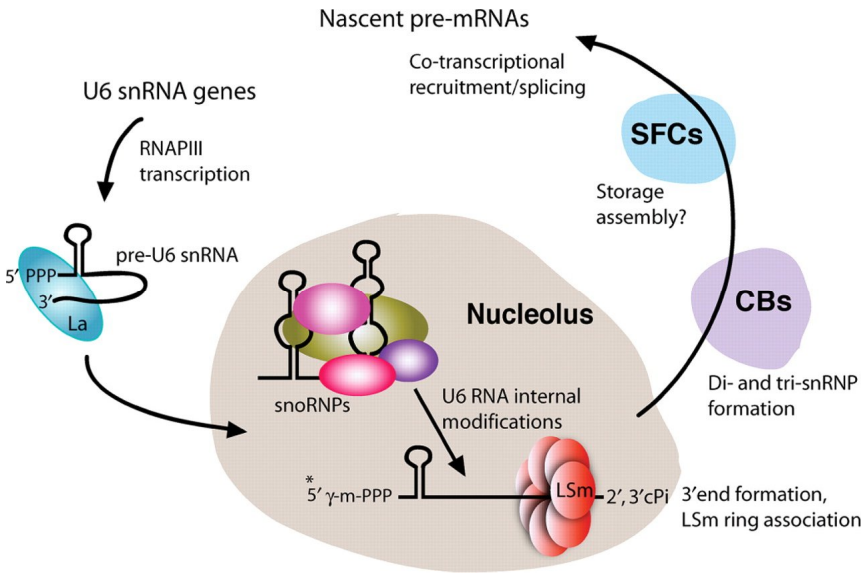


Figure 3.5: The U6 snRNP maturation pathway. Differently from PolII transcribed snRNAs, U6 undergoes a complete maturation cycle in the nucleus, while nucleolus is the place where snRNA modifications are introduced (Patel and Bellini, 2008).

The process of snRNPs maturation is not yet fully understood and a detailed role of some of the factors, e.g. SMN complex, needs to be described. Interesting point is that snRNPs are mobile particles which may be found in distinct nuclear territories and that they associate with storage spaces (nuclear speckles) on one time and later can enter the splicing process in a co-transcriptional manner, thus entering the nucleoplasmic (chromatin) region (Spector and Lamond, 2011).

3.3 Spliceosome assembly

Spliceosome is a highly dynamic RNP which undergoes a series of coordinated reorganizations. These rearrangements mediate the distinct steps of splicing reaction. Spliceosome is known to perform the catalysis of splicing employing snRNP complexes, which assemble *de novo* on the pre-mRNA template during each cycle of splicing. It is widely accepted that spliceosome is assembled in a stepwise fashion and that different snRNPs are associated with the substrate on different time points. In spite of proteomic findings of distinct spliceosomal complexes, a penta-snRNP complex, pre-assembled spliceosome, was also isolated and characterized.

The process of spliceosomal assembly is generally considered as transcription associated and pre-mRNA splicing has been localized to occur in the nucleoplasm, at the site of expression of pre-mRNAs which are about to be spliced. Similarly to other nuclear proteins and RNPs, spliceosomal components seem to be highly dynamic. The existence of chromatin related nuclear environment needs to be taken into account when studying splicing processes. The canonical spliceosome assembly pathway will be described at this point.

3.3.1 Splice sites and BPS recognition (E complex formation)

Splice sites (SS) recognition, the first step of splicing cycle respectively, is of particular importance for this study. It is characterized by E (early) complex formation whose protein composition and internal functional relationships have been extensively studied.

Pre-mRNA molecule possesses several sequence elements which have an impact for the course of splicing (in 5′ to 3′ direction): 5′ splice site (5ss), branch point site (BPS), polypyrimidine tract (PPT) and 3′ splice site (3ss). These elements can be found within an intron and define its borders. Additionally, other regulatory sequences were determined to act in *cis* by recruiting protein factors which are involved in the process of SS recognition: exonic and intronic splicing enhancers and silencers (ESE, ESS, ISE, ISS). The strength of SS is defined according to their complementarity and affinity to spliceosomal components which recognize them, thus U1 snRNA binding for 5ss and affinity to U2AF for 3ss. Generally, mammalian

SS sequences are much less conserved in comparison to highly consensual SS of yeast.

In terms of SS recognition, two major classes of exons can be distinguished: constitutive and alternative. While the constitutive ones always become a part of mature mRNA, the alternative ones, included or excluded, allow several transcripts and peptides expression from a single locus. Alternative splicing is thus responsible for increase of the complexity of proteome (Matlin et al., 2005). Taking into account all the possibilities of regulatory sequences combinations and protein factors involved, exons can be placed on a gradient of exclusion to inclusion with distinct points of constitutive, alternative and predominantly skipped exons (Hertel, 2008).

In vertebrate genome, genes usually contain multiple exons which are significantly shorter that introns. In contrast, intervertebrate (e.g. yeast) genes mostly contain a single short intron (Hawkins, 1988). The length of exons and introns is an important factor for SS recognition. When a large exon is flanked by introns of size smaller than 500 nt, it can be efficiently defined, and vice versa for introns. Nevertheless, recognition of exons and introns is impaired when both elements larger than 500 nt are contained in pre-mRNA at the same time (Sterner et al., 1996). A model of exon definition, where 3´ and 5´ splice sites communicate across an exon, and intron definition, where the SS interaction occurs across an intron, is proposed. The initial recognition of exons and introns thus involves the communication of both splice sites.

The first protein complex formed on pre-mRNA is the H complex. H complex containing mostly heterogeneous nuclear ribonucleoproteins (hnRNPs) appears to be rather unspecific complex which forms on RNA molecule lacking SS and it was shown to assemble *in vitro* (Reed, 1990). First splicing specific ATP independent E complex formation requires 5ss, BPS, PPT and 3ss on the RNA and is composed of several protein factors (see figure 3.6).

In the E complex, U1 snRNA base-pairs with the 5ss and enables its efficient recognition (Michaud and Reed, 1991). The base-pairing itself might not be necessary for 5ss recognition by U1 snRNP as a role of U1C in this step was suggested (Du and Rosbash, 2001, 2002). Mammalian U1 association with pre-mRNA is enhanced by SR proteins. Recently, SR protein SRSF1 (also known as

ASF/SF2) was proposed to bridge the U1 snRNP to the 5ss via U1-70K. The RS domain of SRSF1 needs to be phosphorylated to allow the U1 snRNP association with the E complex (Cho et al., 2011).

The early 3ss (AG dinucleotide) recognition is performed by U2AF35 (Wu et al., 1999), which is a small subunit of U2 auxiliary factor (U2AF). The larger subunit, U2AF65, contains RNA recognition motif and RS domain and is involved in PPT binding in the E complex and subsequent U2 snRNP targeting to this region (Valcarcel et al., 1996; Zamore et al., 1992). U2AF65 plays important role in conformational changes of pre-mRNA molecule and contacts other splicing factors to promote following steps in spliceosome assembly. It bends the RNA molecule so that the PPT and BPS are brought together (Kent et al., 2003). U2AF interaction with pre-mRNA molecule is believed to occur in heterodimers, although existence of U2AF35 homodimers was detected *in vivo* (Chusainow et al., 2005).

Branch point region is recognized by splicing factor SF1 (BBP in yeast) in assistance of U2AF65 recognizing the PPT. The character of interaction where these components cooperate allows recognition of highly degenerate mammalian BPS sequence (Berglund et al., 1998). In contrast, BBP mediated BPS recognition in yeast has a great impact since the yeast branch point sequence is strictly conservative. In mammals, SF1 depletion does not have any dramatic effect on splicing, but still causes some alterations to U2AF65 and U2 snRNP binding (Guth and Valcarcel, 2000).

3.3.2 A complex formation

Pre-catalytic A complex contains U1 and U2 snRNPs and is formed after U2 joins the E complex. Crosslinking analysis confirmed that U1 and U2 snRNAs are stably base-paired with pre-mRNA and almost all U1 and U2 specific proteins are associated with the purified A complex. It is also known that most of proteins identified in the A complex remain associated also with the B complex. Surprisingly, the Prp19/CDC5L complex occurred within the A complex and thus might bind on the splicing complex independently of tri-snRNP (Behzadnia et al., 2007). Prp19/CDC5L complex (NTC – Nineteen complex in yeast) is a particle composed of 7 proteins including Prp19 and CDC5L which form a stable heteromer and are

required for the first catalytic step of splicing. U5 and U6 snRNPs association with the exonic sequences is strengthened by the Prp19/CDC5L or NTC (Makarova et al., 2004). The Prp19 protein is present in 4 copies in the complex (Grote et al., 2010).

The branch point adenosine is strongly committed to the first step of splicing reaction by U2 snRNP binding. The region exhibits a certain level of complementarity to U2 snRNA. A duplex between the snRNA and BPS is thus formed, from which the branch point adenosine is bulged out to form a nucleophile which 2´ OH group is then capable of chemical attack of the 5ss (Query et al., 1994). The U2 snRNP components SF3a and b are proposed to bind the reactive BPS and thus mediate a steric barrier to prevent premature splicing reaction. At a distinct step of spliceosome assembly, Prp2p helicase (a yeast protein) dislocates SF3a and b from the branch point. ATP is required in this step (Lardelli et al., 2010).

The canonical pathway of U2 snRNP joining spliceosome not earlier than in the A complex is partially doubted since U2 components SF3a and b and U2B´´ were found in the E complex. The interaction of U2 snRNP with the E complex is proposed to be weaker than in the A complex and to occur in ATP independent manner (Das et al., 2000). The role of U2 snRNP in functional E complex formation is further supported by the fact that U2 snRNA pseudouridylations are to a certain extend required for proper U1 snRNP recruitment (Donmez et al., 2004). Another evidence for this is the finding that 5´ end of U2 snRNA is localized close to U1 snRNP and both SS already in the E complex. The pre-organized base-pairing of U2 snRNA and BPS is suggested to originate from the E complex interactions (Donmez et al., 2007).

3.3.3 Splice sites cooperation within E and A complex

Splice sites are likely to cooperate in the process of exons and introns definition. This is supported by the finding that they occupy mutually close positions during splicing catalysis. With usage of hydroxyl radical probe tethered to pre-mRNA, it was shown that 5ss, BPS and 3ss appear to be in close proximity in the E complex (Kent and MacMillan, 2002). This idea is further supported in another study, employing similar technique, which revealed that nucleotides bordering the

functional sequences within the intron also lay close to each other (Donmez et al., 2007).

Already in the early time of spliceosome studies, the idea of possible functional SS cooperation in spliceosome assembly was addressed by SS mutations introduction. U1 snRNP was able to bind to the 5ss without ATP requirements and U2 snRNP was then detected in 40S particle on pre-mRNA substrate harboring 3ss mutation, suggesting that U2 snRNP might first contact U1 snRNP before the pre-mRNA substrate. Tri-snRNP was also likely to contact the spliceosomal complexes prior to pre-mRNA binding (Bindereif and Green, 1987). In the absence of either 5ss or 3ss, a sub-fraction of E complex (called E5 or E3) was detected, but presence of both SS increased the effectiveness of assembly. The SS were thus proposed to cooperate in E complex (Michaud and Reed, 1993).

The splicing factors of E complex can be partially redundant when it comes to SS recognition. U1 snRNA depletion from cellular extract results in arrest of splicing, which can be rescued by addition of SR proteins preparation. This set of proteins is capable of performing commitment to the splicing catalysis (Crispino et al., 1994). In *Saccharomyces cerevisiae*, U1 snRNP appears to be associated with 3ss recognition and its compensatory mutations can restore the phenotype of 3ss mutations (Reich et al., 1992). A different situation was described in another yeast organism, *Schisosaccharomyces pombe*, where the compensatory mutations of U1 snRNA were not sufficient to rescue 3ss mutations (Romfo and Wise, 1997). The 3ss itself might not even contribute to efficient spliceosome assembly, as spliceosome was shown to assemble *in vitro* on RNA substrate lacking the 3ss. After addition of molecule containing 3ss YAG trinucleotide (sufficient for 3ss definition) the exons were efficiently ligated (Anderson and Moore, 1997).

The factors which contribute to SS recognition are involved in interactions influencing one another. PPT recognizing U2AF65 was found to help to recruit U1 snRNP to weak 5ss. U2AF65 acts by its RS domain which is capable of binding spliceosomal proteins. It also has a function in stabilization of RNA-RNA interactions. Nevertheless, this effect may be restricted to base-pairing regions which are adjacent to U rich sequences as in case of 5ss (Forch et al., 2003). Vice versa, the character of 5ss can influence the binding of U2AF65. The phenomenon has an impact in alternative exons recognition, where the exchange of 5ss belonging to the

alternative exon for the one of constitutive one, and thus stronger, leads to enhanced U2AF65 binding (Cote et al., 1995). In a presence of weak PPT, there is an interaction of U2AF and U1 snRNP over the exon (Hoffman and Grabowski, 1992). It was also reported that U2AF65 recruitment to the pre-mRNA does not depend on SR proteins interacting with a certain ESE, but requires U1 snRNP (Li and Blencowe, 1999).

BPS is located in the middle of cross intron interactions and influences 5ss and 3ss associated factors. U2AF and U1 snRNP are likely to bind pre-mRNA molecules whose BPS is mutated and form so called E^* complex. E^* to E complex transition is ATP dependent and both U2AF and U1 are loosely associated with E^* complex. The subsequent U2 binding and A complex formation is proposed to require BPS and associated proteins (Champion-Arnaud et al., 1995). Interestingly, in U2AF depleted extract, the so called E´ complex, lacking U2AF but containing U1 snRNP, FBP11 [see (Bedford et al., 1998) for reference] and SF1 is detected. This complex resembles the original E complex organisation. The complex itself is not functional in splicing reaction but can serve as a precursor for E complex formation (Kent et al., 2005).

Some of the E complex interactions were resembled from alternative splicing regulation. Polypyrimidine tract binding protein (PTB) acts on alternative exons exclusion by inhibiting the U2AF binding to the 3ss. As a result of that, the early spliceosomal E complex cannot be assembled. PTB was therein shown to inhibit U2AF binding in 5ss dependent manner, thus excluding U1 binding in the cross exon bridge. The contact of PTB and U1 snRNA was proved (Sharma et al., 2005; Sharma et al., 2011). hnRNP L, another protein of the same group as PTB, works as splicing repressor by blocking the interaction of U1 snRNP or U2AF with SS (Heiner et al., 2010).

Further evidence for SS cooperation is the bridge between 5ss and 3ss factors in mammalian cells mediated by SR proteins. The secondary structure of SR proteins combines an RNA recognition motif (RRM) on the N terminus as well as RS domain (serin and arginin rich) on the C terminus. SR proteins are thus capable of both interaction with target RNA sequences and recruitment of spliceosomal components to adjacent SS (Graveley and Maniatis, 1998). The interactions of SR

proteins which mediate SS communication have a great impact on SS recognition. They were detected in E complex as well as in previously mentioned E5 and E3 complexes but were absent from H complex. SR proteins support the E complex formation in general, while their ESE dependent recruitment contributes to weak SS recognition (Staknis and Reed, 1994). RS domains tethered to ESE were shown to associate with branch point and enhance the E complex assembly (Shen et al., 2004). Additionally, ASF/SF2 and SC35 are known to interact with U1-70K and U2AF35 proteins and are likely to form a bridge between these factors across the intron to enable the SS selection (Wu and Maniatis, 1993). U2AF35 was proposed to act as a bridge in a critical cross exon interaction between U2AF65 and SR proteins bound to U1 snRNP (Zuo and Maniatis, 1996). SC35 is together with U2AF65 required for U2 snRNP recruitment and E to A complex transition. It was also shown to interact with U1 snRNA bound pre-mRNA region and is likely to be absent from this region on molecules lacking functional 3ss (Fu and Maniatis, 1992). In addition, a complex bridging an intron to join the exons together with the assistance of SR proteins was visualized using electron microscopy (Stark et al., 1998).

SR proteins are absent in yeast. Nevertheless, a cross intron network of interacting proteins was detected in yeast as well, proposed to be similar to the mammalian one and containing Prp40 (a component of yeast U1 snRNP), BBP (yeast homologue of SF1) and Mud2p (yeast homologue of U2AF65) (Abovich and Rosbash, 1997).

Vast majority of the results mentioned above was obtained by *in vitro* methods which worked with cell extracts and introduced non-physiological conditions to examine interactions among splicing complexes. There is a certain constriction using these methods as they are less informative when describing the dynamic nature of spliceosome assembly. A different approach is represented by microscopy methods, particularly Fluorescence recovery after photobleaching (FRAP) and Fluorescence resonance energy transfer (FRET). These *in vivo* techniques came into usage in recent years. An example of this approach is a study confirming the interactions of SR proteins ASF/SF2 and SC35 with U1-70K and U2AF35 (Wu and Maniatis, 1993) by FRET. This interaction is indeed present *in vivo* and in addition, SRp20 is also involved in the cross exon bridge. Striking are the findings of spatial relationships between SR proteins and splicing factors: they

accumulate in nuclear speckles and their interaction is weakened in the nucleoplasm upon transcription inhibition (Ellis et al., 2008). This brings forward the idea of preformed complexes of splicing factors. Such complexes containing both U2AF subunits and SF1 protein were detected by co-immunoprecipitation. The existence of preformed complexes was also examined by FRAP experiments, suggesting lower mobility of U2AF and SF1 and existence of both transient and stable interactions among them (Rino et al., 2008).

The previously discussed concept of SS communication within E complex is extended to the phase of assembly where SF1 and U2AF are displaced from the spliceosome and substituted with U2 snRNP, thus at the stage of A complex. U2 snRNP has been expected to somehow interact with U1 snRNP across the intron or the exon. In *Schizosaccharomyces pombe*, Prp5 is likely to bridge U1 and U2 in ATP dependent manner. Strikingly, the two snRNPs were found to form a single particle, U1-U2 di-snRNP. Prp5 was shown to interact independently with both of the snRNPs and involved domains were identified. In the absence of U1 binding domain, the activity of U2 is significantly decreased which confirms the functional relationship between U1 and U2 snRNP. It is also important to notice that Prp5 associates with U1 and U2 snRNPs in the absence of pre-mRNA molecule (Xu et al., 2004). In addition, there is a pathway of SR proteins dependent assembly of U2 which works in cooperation with U1-70K protein. This U1 snRNP specific factor is bound by the ESE associated SR proteins which also interact with U2AF and SF1 over the upstream exon (Boukis et al., 2004).

Indeed, U1 snRNP recruitment is significantly decreased on a gene carrying 5ss mutation. In absence of U1:5ss base-pairing, the U2 snRNP recruitment is also compromised, further suggesting the cooperative character of spliceosome assembly, where U1 snRNP creates an interface for rest of the snRNPs to bind (Lacadie and Rosbash, 2005).

In contrast to previously mentioned SS cooperation and interactions occurring in the E complex, there is a study suggesting that the SS pairing requires ATP and takes place not earlier than in the A complex, containing stably bound U2 snRNP (Kotlajich et al., 2009).

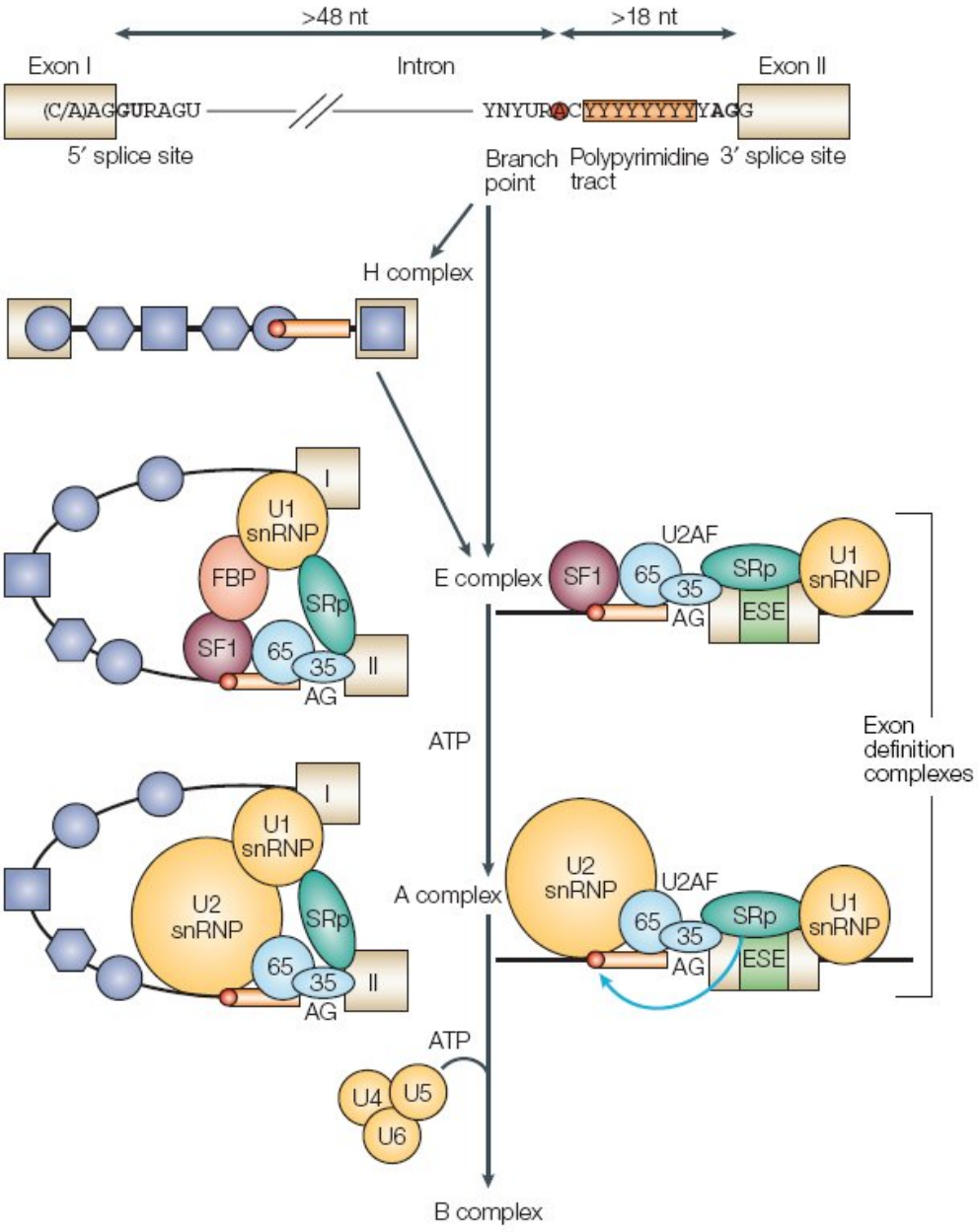


Figure 3.6: The early steps of spliceosome assembly. In the top of the picture, the most important sequence elements within an intron are depicted. Bellow that, a scheme of spliceosome assembly in the early steps is shown. H complex presents the phase of transcription associated recruitment of hnRNPs. E, A and B complex are assembled afterwards. Both intron and exon definition assembly pathways are depicted. SRp – SR protein, SF1 – splicing factor 1, FBP – forming binding protein, ESE – exonic splicing enhancer, 65, 35 – U2AF65, 35 (Matlin et al., 2005).

3.3.4 Tri-snRNP association (B and C complex formation)

Further step in spliceosome assembly is B and C complex formation which is characterized by extensive structural rearrangements and which leads to catalysis of the two transesterification reactions. Tri-snRNP joins spliceosome by that time.

B complex comprises all spliceosomal snRNA molecules. In addition to proteins contained already in the A complex, a complete list of tri-snRNP proteins and a large number of non-spliceosomal factors involved in splicing was detected in B complex. This abundant group of proteins thus joins spliceosome before the catalytic step occurs. Important point is that Prp5, a protein with a role in U1-U2 snRNPs bridging in A complex (Xu et al., 2004), is absent in B complex. This indicates a possibility that the event of U1 and U2 contact is only mediated in early stages of splicing and is not retained for the rest of splicing cycle. B complex additionally contains for example U2AF65 and U2AF35 and a set of proteins of Exon junction complex (EJC, proteins deposited 20 nucleotides upstream of exon junctions) but does not contain some of the key factors of splicing catalysis (Deckert et al., 2006).

During B to C complex transition, broad conformational changes take place and proteomic content of spliceosome is significantly altered. Some of the proteins which were only weakly associated with the spliceosome become tightly bound and thus able to perform the catalysis. There are certain stages between B and C complex, called B^{act} and B^{*}, whose composition reflect the transformation from pre-catalytic to catalytically active spliceosome which performs the splicing reactions (Bessonov et al., 2010). This transition is particularly characterized by U1 and U4 snRNPs disassembly from the spliceosomal B complex. B^{*} complex is the one to perform the first transesterification reaction involving BPS nucleophilic attack on the 5ss which results in excision of the 5′ end of an intron (Wahl et al., 2009).

Finally, C complex carries out the second step in which the intron is released in form of lariat. Structural repositioning of spliceosomal core allows the 3ss to enter the active site. The key feature of second step is disturbing of 5ss-U6 snRNA binding. This interaction, in addition to U2-U6 snRNAs base-pairing, is important for the first catalytic step but inhibits the second one (Konarska et al., 2006). In C

complex, all U4/U6 di-snRNP proteins, including U6 Lsm proteins, are released from the splicing machinery. In absence of other factors, C complex is capable of 2 exons ligation, thus retaining the most important catalytic functions. In terms of protein composition, several C complex specific proteins were identified, e.g. splicing factors Slu7 and Prp22 and helicases Abstract and DDX35. Prp19 and associated factors are more enriched in C complex, while SF3a and b are assumed to be less abundant (Bessonov et al., 2008). The highly stable catalytic core of C complex exhibits a significant similarity to postspliceosomal 35S U5 snRNP particle. It contains all of the important factors of U5 origin, e.g. Prp8, Snu114 and U5 snRNA (Golas et al., 2010).

At this point, an alternative view on tri-snRNP assembly to the spliceosomal machine can be taken. In a traditional assembly pathway, tri-snRNP joins pre-mRNA at the stage of A complex to form B complex. In spite of this, several studies proved that there might be some alternatives. *In vitro* evidence for these considerations is 5ss carrying oligonucleotide which is capable of tri-snRNP recruitment in the absence of U2 snRNP. This step is enhanced by blocking of U1 snRNA binding to the oligonucleotide, the interruption of U1-5ss interaction is thus a key step in proceeding though the assembly pathway. After U2 addition to the system, the tri-snRNP associated oligonucleotide enables its association to the entire complex (Konforti and Konarska, 1994).

Tri-snRNP association with 5ss was confirmed by another study which suggested this event to be independent of U2 snRNP binding to the BPS. Altered mechanism of 5ss selection might include U1 snRNP and tri-snRNP simultaneous binding while tri-snRNP would be responsible for increase in specificity of 5ss recognition. The interaction required ATP as tri-snRNP was discussed to take a certain ATP dependent conformation. A direct interaction between 5ss and U6 snRNA was detected (Maroney et al., 2000). In a trans-splicing *in vitro* system, U6 and U4 are likely to contact 5ss in ATP independent manner and even in absence of U1, probably acting as precursors for following tri-snRNP recruitment. The U4/U6 interaction is likely to be converted into ATP dependent one in later stages of spliceosome assembly (Johnson and Abelson, 2001).

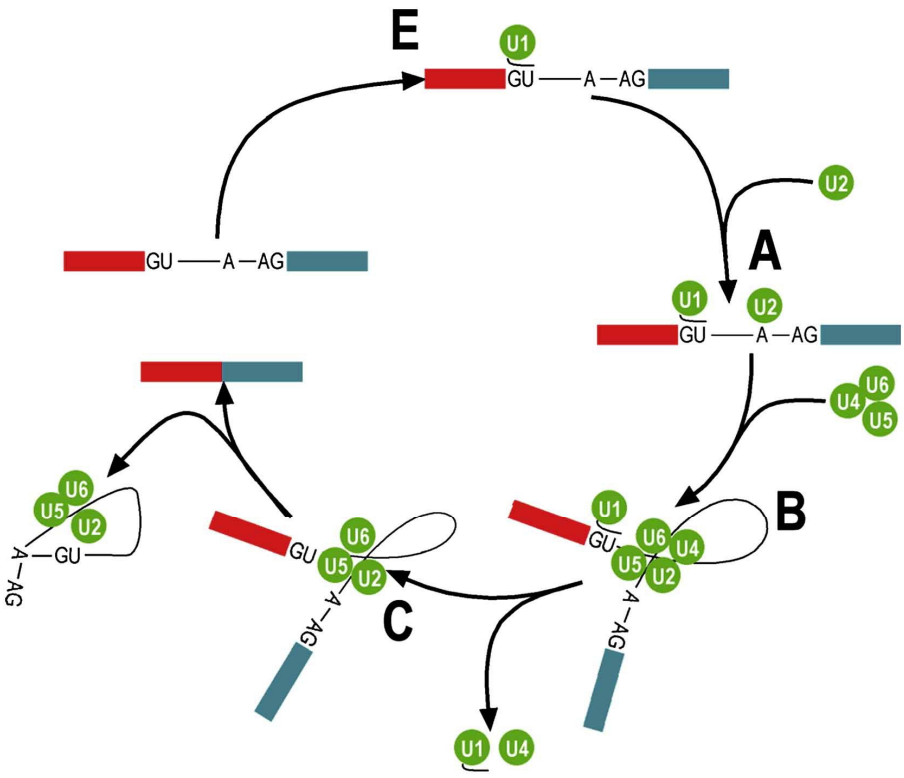


Figure 3.7: A schematic view of canonical spliceosomal stepwise assembly pathway. In E complex, U1 snRNP associates with 5ss. U2 snRNP is likely to join U1 in the A complex. Tri-snRNP containing U4, U5 and U6 snRNPs is recruited afterwards to create B complex which than catalyzes the first transesterification of splicing reaction. U1 and U4 leave the B complex and the rest forms C complex. The second transesterification is then catalyzed, intron is released in the form of lariat and exons are joined together in a mature mRNA molecule (Sperling et al., 2008).

The idea of exon and intron definition complexes was discussed in previous chapters. In a recent study, an exon which acts as a scaffold for exon definition complexes was used to resemble the pathway of tri-snRNP in spliceosome assembly. The upstream intron was 5ss-less and upon short 5ss containing oligonucleotide addition, a trans-splicing system was created for mature mRNA production. In such a system, tri-snRNP was observed to contact the exon via U2 snRNA – U6 snRNA di-helix and a potential set of protein interactions. U2 was necessary for tri-snRNP recruitment. When the 5ss oligonucleotide was added to the

system, the tri-snRNP interaction with the exon complex was stabilized and spliceosomal B complexes were formed, suggesting a possible role of tri-snRNP in transition from cross exon to cross intron bridge in pre-mRNA utilization (Schneider et al., 2010).

3.3.5 Penta-snRNP versus stepwise assembly

As a challenge for canonical stepwise pathway of splicing complex formation, yeast penta-snRNP containing all five spliceosomal particles was efficiently purified using affinity chromatography and characterized (Stevens et al., 2002). The penta-snRNP itself was not capable of splicing catalysis. Nonetheless, after nuclease treated yeast nuclear extract addition, the catalysis was enabled. The isolated complex was shown not to contain any pre-mRNA molecule or pre-mRNA associated protein (e.g. Cap binding complex or hnRNP). It also did not contain a set of splicing associated proteins, particularly ATPases and splicing factors required for catalytic steps. The penta-snRNP mediated splicing catalysis was suggested to run in a way similar to stepwise model, employing the very same components with the only difference that those factors are likely to be assembled in the preformed complex during the catalysis (Stevens et al., 2002). Penta-snRNP was detected to assemble also in mammalian cells *in vitro*, on a 5ss carrying oligonucleotide. The complex recognized 5ss via U1 snRNP in the first step, while U5 occupied a place of U1 in the second step. Spliceosome formed a single large particle during the catalysis (Malca et al., 2003).

Supraspliceosome model presents the idea of multiple spliceosomes assembled on pre-mRNA molecule in groups of four. These large molecular machines, detected by electron microscopy and purified, would enable a much more efficient SS recognition, increasing the spliceosome fidelity, by recognizing several introns at one time. Suprasoliceosomes are likely to be reconstituted *in vitro* in presence of pre-mRNA molecule (Azubel et al., 2006).

With respect to recently used *in vivo* approaches, the findings of penta-snRNP have been examined by various studies. Chromatin immunoprecipitation (ChIP) technique allows partial insight into transcription and splicing relationships. Using ChIP, different recruitment of snRNPs at different time

was shown (Gornemann et al., 2005). This idea was proposed for yeast cells where U1 snRNP is recruited to the site of transcription and thus is cross-liked to the chromatin prior to U2 snRNP and tri-snRNP recruitment. This assembly pathway is also in agreement with a "timescale" of snRNPs recruitment along the gene (closer to 5ss means earlier). snRNPs are likely to associate with the transcriptional locus in a stepwise fashion. The Cap binding complex, a set of proteins contacting the 5´ 7-methyl guanosine cap, is in case of some genes responsible for snRNPs recruitment to the transcription unit (Gornemann et al., 2005). It is assumed from *in vivo* depletion and ChIP experiments that U1 is required for binding of U2 and tri-snRNP. In contrast, U1 is recruited in U2 independent manner. Upon U5 depletion, U1 and U2 containing complexes were detected. These results are in contrast with the idea of penta-snRNP as well as U1 plus tetra-snRNP pathways previously considered to work in yeast (Tardiff and Rosbash, 2006).

Spliceosome assembly inhibition by Spliceostatin A, which targets the U2 snRNP component SF3b, revealed that at the stage of inhibition, U1 and U2 snRNPs are stably associated with the pre-mRNA molecule. Tri-snRNP is not stably associated with the complex, although it may slightly interact with it (Roybal and Jurica, 2010). *In vitro* evidence for stepwise assembly is the structural approach: in tri-snRNP depleted cell lysate, thus under conditions which do not allow penta-snRNP formation, stable and functional A complex forms which is than capable of splicing catalysis after necessary factors addition (Behzadnia et al., 2006).

Employing a novel microscopy technique, spliceosome stepwise assembly was recently confirmed, although with a usage of cellular extract and thus not bringing complete information (Hoskins et al., 2011). As already mentioned, microscopic methods based on live cell imaging enable closer description of the dynamic nature of nuclear proteins. Based on importance of *in vivo* conditions, a recent study of snRNPs dynamics revealed a differential time of interaction of single snRNPs with pre-mRNA, using FRAP and Fluorescence correlation spectroscopy (FCS) methods. That would clearly speak for stepwise assembly of spliceosome. In addition, the first estimate of how long splicing reaction takes was brought (Huranova et al., 2010).

4. Materials and methods

4.1 Materials

4.1.1 List of instruments

Leica TCS SP5 Confocal Microscope (Leica)

DeltaVision (Applied Precision)

CO_2 incubator (Jouan)

Flow box (Jouan)

Gel Dox XR+ (BioRad)

Horizontal electrophoresis (BioRad)

Vertical electrophoresis (BioRad)

Electrophoresis for RNA gels

Trans-Blot SD Semi-Dry Transfer Cell (BioRad)

Odyssey Infrared Imaging System (LI-COR)

LightCycler 480 System (Roche Applied Science)

PCR cycler MJ Mini Gradient (BioRad)

Centrifuge (Biofuge)

Centrifuge (Eppendorf)

Rocker (Biosan)

Vortex (Scientific Industries)

Rotator (Biosan)

Sonicator (Kika Labortechnik)

Thermo shaker (Biosan)

Shaker 37°C (Lab. Companion)

ND-1000 Spectrophotometer (NanoDrop)

Power supplies (BioRad)

4.1.2 Cell lines

HeLa cells were cultured in high-glucose DMEM (4.5 g/l glucose) supplemented with 10% fetal calf serum, penicillin, and streptomycin (Invitrogen). MS2-RFP stable cell line was generated by transfection of MS2-RFP construct with

Fugene (Roche Applied Science) and selection with G418 (final concentration 1 mg/ml) for one week. E3 stable cell lines were generated with usage of MS2-RFP stable cell line which was co-transfected with E3 vector and puromycin resistence encoding vector pLPCX (Clontech). After approximately ten days of puromycin selection (final concentration 625 ng/ml), positive colonies were picked from the plate and grown separately. The microscopy pattern of individual clones was then analyzed by DeltaVision fluorescent microscope (Applied Precision).

4.1.3 Bacterial strains

We used One Shot Max Efficiency DH5α-T1R competent bacteria (Invitrogen) with genotype: F- Δ(*lac*ZYA-*arg*F)U169 *rec*A1 *end*A1 *hsd*R17(rk-,mk+)*pho*A *sup*E44*thi*-1 *gyr*A96 *rel*A1 *ton*A. Bacteria were grown in sterile conditions using LB medium. Ampicilin (50 μg/ml) was used for selection of transformed bacteria.

4.1.4 Antibodies

- anti-GFP goat antibody used for immunoprecipitation was provided by David Drechsler (Max Plank Institute, Dresden).
- anti-GFP mouse antibody (Santa Cruz) was used as primary antibody for Western blots.
- anti-Sm (Y12) mouse antibody used as primary antibody for Western blots was produced by Institute of Molecular genetics, Academy of Sciences of the Czech Republic core facility (Stanek et al., 2008)
- anti-mouse IRDye 800CW donkey antibody (LI-COR) was used as secondary antibody for Western blots.

4.1.5 Plasmids

E3 vector containing β-globin gene was a gift from Yaron Shav-Tal (Bar-Ilan University, Israel). The original vector [see (Darzacq et al., 2006)] is based on pcDNA3 vector backbone (Invitrogen) with deleted G418 resistance gene and

contains minimal CMV promoter under control of tetracycline response element. We exchanged this promoter for regular CMV promoter from pcDNA3 vector using NruI and HindIII restriction sites (see figure 5.1).

To generate E3_del-in1-WT and E3_del-in2-WT vectors, we deleted nucleotides 143 – 272 and 496 – 1536 of β-globin coding sequence contained in the vector by PCR (see figure 5.2). We used primers with 5´ phosphate group to enable blunt ends ligation. E3_del-in1-M1 and M2 as well as E3_del-in2-M1 and M2 were generated by site directed mutagenesis. One (M1) or two (M2) mismatches were introduced during PCR reaction with mismatching primers (see figure 5.2 for exact sequence of mutations).

pRed-bGlo_5M, pRed-bGlo_3m5M, pRed-bGlo_3M and pRed-bGlo_5m3M were a gift from Massimo Caputi (University of Florida, USA) (see figure 5.11). The constructs are based on pHcRed1-C1 vector (Takara Bio Inc.) and contain a single MS2 stem loop and the following segment of β-globin gene:

GAGGCCCTGGGCAGgttggtatcaaggttacaagacaggtttaaggagaccaatagaaactgggcatgtgg
agacagagaagactcttgggtttctgataggcactgactctctctgcctattggtctattttcccacccttagGCTGCTG
GTGG

(Capitals indicate contained pieces of exon 1 and 2 of β-globin gene). pRed-bGlo_3m5M and pRed-bGlo_5m3M harbor 3ss respectively 5ss mutation (AG>TT, GT >AA).

pRed-bGlo_5m5M and pRed-bGlo_3m3M were generated by site directed mutagenesis, introducing two mismatches to the first two or the last two nucleotides of the β-globin intron 1. GT was changed for AA and AG for TT (see figure 5.12 for complete scheme of pRed-bGlo vectors).

MS2-RFP, MS2-YFP, U1C-CFP and U2B´´-GFP vectors were cloned by Martina Huranová (RNA biology department, Institute of Molecular genetics, Academy of Sciences of the Czech Republic) and are based on pDsRed-Monomer-N1, pEYFP-N1, pECFP-N1 and pEGFP-N1 vectors (Takara Bio Inc.) [see (Huranova et al., 2010) for MS2-RFP and U2B´´-GFP]. U2AF35-CFP was provided by M. Caputi and is based on pCerulean-C1 vector (modified pECFP-N1, Takara Bio Inc.).

4.1.6 Primers

Table 4.1: A list of primers

Name of the primer	Sequence (5′ to 3′ direction)	Used for
E3delin1_F	GCTGCTGGTGGTCTACCCTTG	β-globin intron 1 deletion
E3delin1_R	CTGCCCAGGGCCTCACCACCA	β-globin intron 1 deletion
E3delin2_F	CTCCTGGGCAACGTGCTGGTC	β-globin intron 2 deletion
E3delin2_R	CCTGAAGTTCTCAGGATCCAC	β-globin intron 2 deletion
E3in2M1_F	CTGAGAACTTCAGGGGGAGTCTATGGGACGC	Mutagenesis of 5ss of β-globin intron 2
E3in2M1_R	GCGTCCCATAGACTCCCCCTGAAGTTCTCAG	Mutagenesis of 5ss of β-globin intron 2
E3in2M2_F	CCTGAGAACTTCAGGGGGAATCTATGGGACGCTTGA	Mutagenesis of 5ss of β-globin intron 2
E3in2M2_R	TCAAGCGTCCCATAGATTCCCCCTGAAGTTCTCAGG	Mutagenesis of 5ss of β-globin intron 2
E3in1M1_F	GAGGCCCTGGGCAGGGTGGTATCAAGGTTAC	Mutagenesis of 5ss of β-globin intron 1
E3in1M1_R	GTAACCTTGATACCACCCTGCCCAGGGCCTC	Mutagenesis of 5ss of β-globin intron 1
E3in1M2_F	GTGAGGCCCTGGGCAGGGTGATATCAAGGTTACAAGAC	Mutagenesis of 5ss of β-globin intron 1
E3in1M2_R	GTCTTGTAACCTTGATATCACCCTGCCCAGGGCCTCAC	Mutagenesis of 5ss of β-globin intron 1
E3del-in1-mRNA_F	TTGCCACACTGAGTGAGCTG	E3_del-in1 splicing products detection, qPCR
E3del-in1-mRNA_R	CACACAGACCAGCACGTTG	E3_del-in1 splicing products detection, qPCR
E3del-in1-pre-mRNA_F	GCTCACCTGGACAACCTCA	E3_del-in1 splicing products detection, qPCR
E3del-in1-pre-mRNA_R	GGGGAAAGAAAACATCAAGC	E3_del-in1 splicing products detection, qPCR
E3del-in2-mRNA_F	CAAGGTGGACGTGGATGAAG	E3_del-in2 splicing products detection, qPCR
E3del-in2-mRNA_R	GGACAGATCCCCAAAGGACT	E3_del-in2 splicing products detection, qPCR
E3del-in2-pre-mRNA_F	CAAGGTGGACGTGGATGAAG	E3_del-in2 splicing products detection, qPCR
E3del-in2-pre-mRNA_R	GTCTCCACATGCCCAGTTTC	E3_del-in2 splicing products detection, qPCR
E3totalRNA_F	TTGGACCCAGAGGTTCTTTG	qPCR primer in RIP experiments, E3 total RNA detection
E3totalRNA_R	CTTTCTTGCCATGAGCCTTC	qPCR primer in RIP experiments, E3 total RNA detection
5M-mRNA_F	CGGACTCAGATCTCGAGGAG	pRed-bGlo_5M splicing products detection
5M-mRNA_R	TCAGTTATCTAGATCCGGTGGA	pRed-bGlo_5M splicing products detection
5M-pre-mRNA_F	CGGACTCAGATCTCGAGGAG	pRed-bGlo_5M splicing products detection
5M-pre-mRNA_R	GTCTCCACATGCCCAGTTTC	pRed-bGlo_5M splicing products detection

3M-mRNA_F	CAAGAGGATTACCCTTGTTCG	pRed-bGlo_3M splicing products detection
3M-mRNA_R	CAAATGTGGTATGGCTGATTATG	pRed-bGlo_3M splicing products detection
3M-pre-mRNA_F	CAAGAGGATTACCCTTGTTCG	pRed-bGlo_3M splicing products detection
3M-pre-mRNA_R	TTGGTCTCCTTAAACCTGTCTTG	pRed-bGlo_3M splicing products detection
5m5M_F	GAGGAGGCCCTGGGCAGAATGGTATCAAGGTTACAA	pRed-bGlo_5M 5ss mutagenesis
5m5M_R	TTGTAACCTTGATACCATTCTGCCCAGGGCCTCCTC	pRed-bGlo_5M 5ss mutagenesis
3m3M_F	GGTCTATTTTCCCACCCTTTTGCTGCTGGTGGGGATC	pRed-bGlo_3M 3ss mutagenesis
3m3M_R	GATCCCCACCAGCAGCAAAAGGGTGGGAAAATAGACC	pRed-bGlo_3M 3ss mutagenesis
EGFP-N-seq	CGCCGTCCAGCTCGACCA	Primer for reverse transcription

4.2 Methods

4.2.1 Restriction digestions and ligation

Restriction digestions were performed with usage of appropriate restriction enzymes (New England BioLabs) at prescribed temperatures and in prescribed buffer, eventually with bovine serum albumin (BSA) supplement. The restriction enzymes were inactivated at prescribed temperatures. When performing sequential digestions, reactions were purified with QIAquick PCR Purification Kit (Qiagen) between single digestions. The products of restriction digestions were analyzed on 1% agarose gel and bands of interest were cut out. DNA was purified by QIAquick Gel Extraction Kit (Qiagen). The extracted DNA was analyzed on 1% agarose gel again to determine concentrations of vector and insert. In the subsequent ligation reaction, 3x higher amount of insert than vector was used. T4 DNA ligase (Fermentas) was used for the reaction run in 16 °C overnight. The ligation reaction was transformed into competent bacteria on the next day.

Table 4.2: Restriction digestion

Enzyme1	1 µl
Enzyme2	1 µl
10x conc. buffer	2 µl
DNA	1-2 µg
Nuclease free H_2O	to 20 µl

Table 4.3: Ligation reaction

T4 ligase	1 µl
10x conc. ligation buffer	2 µl
Molar ratio insert:vector	3:1
Nuclease free H_2O	to 20 µl

4.2.2 Transformation of competent bacteria

Competent strain of DH5α bacteria was used for ligation products transformation. Bacteria were thawed on ice and 100 µl was used for one transformation. Ligation reaction added to the bacterial suspension never exceeded 10 % of its volume. After 20 minutes of incubation on ice, the mixture was exposed to 35 seconds heat shock at 42 °C. 400 µl of SOC medium was added and the mixture was incubated at 37 °C for 1 hour and subsequently grown on agar plate containing antibiotics at 37 °C overnight.

Table 4.4: SOC medium composition (100 ml)

2 g tryptone
1 ml 2M glucose
0.25 ml 1M KCl
1 ml 1M NaCl
1 ml 1M $MgSO_4$
1 ml 1M $MgCl_2$
0.5 g yeast extract

4.2.3 Plasmid DNA preparation from bacterial culture

Colonies picked from the agar plate were grown in 2 ml of LB medium (1% tryptone, 0.5% yeast extract, 1% NaCl) with ampicilin in shaker at 37 °C overnight. On the next day, 1.5 ml of bacterial culture was used for plasmid DNA isolation by QIAprep Spin Miniprep Kit (Qiagen). DNA was eluted by 50 µl of elution buffer and concentration was measured by ND-1000 Spectrophotometer (NanoDrop). Products of ligation were checked by restriction analysis.

4.2.4 Transfection of HeLa cells

HeLa cells were transfected using Fugene HD transfection reagent (Roche Applied Science). A Fugene (volume):DNA (weight) ratio 3:1 was determined as the most efficient one. After mixing Fugene, DNA and serum free medium, the mixture was incubated for 15 minutes in room temperature and added to the cells in droplets manner. Cells were used for experiments 18 – 24 hours after transfection.

Table 4.5: Composition of transfection mixtures

	Fugene	DNA	medium
6 well plate (single well)	6 µl	2 µg	120 µl
10 cm Petri dish	24 µl	8 µg	500 µl

4.2.5 Site directed mutagenesis

Mutagenesis was performed by PCR with mismatching primers designed by QuickChange Primer Design program (Stratagene, available online at http://www.stratagene.com/qcprimerdesign). Primers are listed in table 4.1. Templates were amplified by Phusion DNA polymerase (Finnzymes). PCR products were purified by QIAquick PCR purification kit (Qiagen) and digested with DpnI restriction endonuclease (New England BioLabs) at 37 °C for 2 hours to get rid of residual templates. The mixture was then transformed into competent bacteria and on the next day, individual colonies were picked for plasmid DNA preparation.

Table 4.6: PCR reaction for site directed mutagenesis

5x conc. Phusion GC buffer	10 µl
2% DMSO	10 µl
dNTPs (10mM each)	1 µl
Phusion polymerase (2 U/µl)	0.5 µl
primer F (10µM)	0.5 µl
primer R (10µM)	0.5 µl
DNA template	50 ng
H_2O	to 50 µl

Table 4.7: PCR program for site directed mutagenesis

1.	98 °C	30 seconds
2.	98 °C	10 seconds
3.	55 °C	30 seconds
4.	72 °C	30 sec/1 kb of template
5.	Go to 2	24 times
6.	72 °C	10 minutes
7.	4 °C	forever

4.2.6 Total RNA isolation

RNA from HeLa cells was isolated using TRI Reagent (Applied Biosystems). 500 µl of TRI Reagent per single well of 6 well plate were used to lyse the cells. After 5 minutes of room temperature incubation and careful mixing, cell lysate was transferred to Eppendorf tube and 100 µl of chloroform was added. The mixture was vortexed and centrifuged at 14 000 rpm and 4 °C for 15 minutes. The aqueous phase was transferred to clean Eppendorf tube and nucleic acids were precipitated upon addition of 250 µl of isopropanol at room temperature for 15 minutes. The samples were centrifuged at 14 000 rpm and 4 °C for 10 minutes, supernatants were discarded and pellets were washed with 500 µl of 70% ethanol. Dried pellets were resuspended in 50 µl of nuclease free water. Quality of isolated RNA was checked on 1% agarose gel and concentration was measured on ND-1000 spectrophotometer (NanoDrop).

4.2.7 DNAse treatment

Izolated RNA was treated with rDNAse 1 (DNAfree kit, Ambion). When total cellular RNA was isolated, 2 µg of RNA were used for the reaction. 5 µl out of 15 µl were used when RNA isolated from IP preparations/inputs (RIP) was treated. RNA was treated at 37 °C for 30 minutes. 2 µl of DNAse inactivation reagent were added afterwards and mixed with the sample for 2 minutes. Samples were centrifuged at 14 000 rpm for 2 minutes. DNA free RNA was then added to reverse transcription reaction.

Table 4.8: DNAse treatment reaction mixture

RNA	2 µg/5 µl
10x conc. buffer	1 µl
rDNAse 1	1 µl
H_2O	to 10 µl

4.2.8 Reverse transcription

Reverse transcription (RT) of isolated RNA was performed using SuperScript III Reverse Transcriptase (Invitrogen). Complete procedure was performed on ice. 1 µg of DNAse treated total cellular RNA (or 8 µl of DNAse treatment reaction, when working with IP samples/inputs from RIP) was used for RT. After mixing of RNA, RT primer, deoxynucleotide triphophates (dNTPs, Fermentas) and water, reaction mixture was incubated at 65 °C for 5 minutes. Reverse transcriptase, 0.1M DTT and buffer were added afterwards and reaction ran at 55 °C for 30 - 45 minutes. Reverse transcriptase was inactivated in 70 °C for 15 minutes.

Table 4.9: RT reaction mixture

RNA	1 µg/8µl
RT primer (2 pmol/µl)	1 µl
dNTPs (10mM each)	1 µl
5x conc. First Strand buffer	4 µl
DTT (0,1M)	1 µl
SuperScript Reverse Transcriptase	0.5 µl
H_2O	to 20 µl

4.2.9 RT-PCR

RT produced cDNA was amplified by PCR using Taq polymerase (Fermentas). Primers are listed in table 4.1.

Table 4.10: PCR reaction for RT-PCR

cDNA	2 µl
10x conc. buffer	2.5 µl
dNTPs (10mM each)	0.5 µl
$MgCl_2$ (25mM)	1.5 µl
Primer F (100µM)	0.25 µl
Primer R (100µM)	0.25 µl
Taq polymerase	0.5 µl
H_2O	to 25 µl

Table 4.11: PCR program for RT-PCR

1.	95 °C	4 minutes
2.	95 °C	2 minutes
3.	55 °C	30 seconds
4.	72 °C	1 minute
5.	Go to 2	24 times
6.	72 °C	5 minutes
7.	4 °C	forever

4.2.10 Quantitative PCR (qPCR)

qPCR was used for amplification of cDNA from both total cellular RNA preparations and IP preparations/inputs (RIP). We used LightCycler 480 System (Roche Applied Science). When working with total RNA samples, cDNA was diluted 10 times with water before qPCR was performed. IP/input samples from RIP were not diluted. No-RT control was included to make sure that there is no genomic DNA contamination. Each sample was pipetted in duplicate. The reaction mixture was composed of 0.5 µl of 5µM primers (forward + reverse), 2.5 µl of SybrGreen

(Roche Applied Science) and 2 µl of cDNA (5 µl total). Primers are listed in table 4.1.

Normalization of qPCR of total RNA preparations (splicing efficiency experiments, figures 5.5 and 5.6) was performed in the following way:

relative mRNA abundance = $1/2^{[[Ct(mRNA)-Ct(totalRNA)]mut -[Ct(mRNA)-Ct(totalRNA)]wt]}$

relative pre-mRNA abundance = $1/2^{[[Ct(pre\text{-}mRNA)-Ct(totalRNA)]mut - [Ct(pre\text{-}mRNA)-Ct(totalRNA)]wt]}$

Normalization qPCR for RIP experiments (analysis of co-purified RNA, figures 5.8 and 5.21) was normalized on input levels and on background in the following way:

fold enrichment over background = $2^{[Ct(input)-Ct(IP)]specific}/2^{[Ct(input)-Ct(IP)]unspecific}$

4.2.11 Cell lysate preparation (for RNA immunoprecipitation)

Cells were grown in 10 cm Petri dishes and transfected with combinations of E3 vectors (two times higher amount of E3_del-in1-M2 than E3_del-in1-WT) and U2B´´-GFP 24 hours prior cell lysate preparation. Cells from each well were washed 3 times with PBS and scraped into 1 ml of ice-cold PBS. The suspension was centrifuged at 500 g and 4 °C for 5 minutes and pellets were resuspended in 1 ml of RIP lysis buffer A supplemented with 5 µl of Protease inhibitors III cocktail (Calbiochem) and 2 µl of RNAsin (Promega). Samples were vortexed and sonicated on 30% performance by 30 pulses on ice. The lysed nuclei were checked on microscope. Samples were then centrifuged at 20 000 g and 4 °C for 10 minutes. Supernatants were used for IP. 40 µl of lysate was used as input.

4.2.12 Immunoprecipitation (for RNA immunoprecipitation)

Immuniprecipitation (IP) with Protein-G Sepharose beads (GE Healthcare) was performed with goat α-GFP antibody. Antibody (0.4 µl, 15 mg/ml) was bound on 30 µl of 3 times washed beads in 450 µl of RIP lysis buffer A. Blocking of beads with 200 µg/ml of yeast tRNA (Ambion) and 500 µg/ml of BSA (New England BioLabs) was performed at the same time. Beads with antibody and blocking reagents were incubated on rotator at 4 °C for 2 hours. Afterwards, beads were washed for 3 times with RIP lysis buffer A and cell lysate was added to them.

Beads with cell lysate were incubated on rotator at 4 °C for 4 hours. In the end, beads were washed for 5 times with RIP lysis buffer B and saved at -80 °C.

Table 4.12: RIP lysis buffer A and B

Buffer A	Buffer B
20mM Tris-HCl pH 7.4	20mM Tris HCl pH 7.4
0,2M KCl	0,2M KCl
2mM $MgCl_2$	2mM $MgCl_2$
1% (v/v) NP-40	0.5% (v/v) NP-40

4.2.13 RNA isolation from sepharose beads

10 µl of sepharose beads from IP (out of 30 µl) and 10 µl of input samples were mixed with 90 µl of RIP lysis buffer B, 10 µl of 10% SDS and 5 µl of glycerol. Samples were vortexed and 200 µl of phenol-chlorophorm-isoamylalcohol was added. Samples were then incubated shaking at 37 °C for 15 minutes and centrifuged at 20 000 g and 4 °C for 5 minutes. Water phase was transferred to clean Eppendorf tube. 12 µl of 3M sodium acetate and 500 µl of ~100% ethanol were added. Nucleic acids were precipitated at -20 °C for 30 minutes. Samples were then centrifuged at 20 000 g and 4 °C for 10 minutes. Pellets were washed with 70% ethanol, dried and resuspended in 15 µl of nuclease free water. Subsequent DNAse treatment, RT and qPCR analysis are described in chapters 4.2.7, 4.2.8 and 4.2.10.

4.2.14 SDS PAGE

20 µl of beads from IP preparation and 30 µl of input samples were mixed with equal volume of 2x concentrated sample buffer (prepared from 5x conc.) and incubated at 95 °C for 5 minutes. Samples were then loaded on polymerized 5% stacking gel poured on separating gel (12% gel for MS2-YFP IP, 10% gel for U2B´´-GFP IP) between glasses in gel chamber (BioRad). Electrophoresis ran at 45 V for 20 minutes and then at 120 V for 90 minutes in SDS PAGE running buffer.

Table 4.13: 5x concentrated Sample buffer

10% (w/v) SDS
10mM DTT
20% (v/v) Glycerol
0.2M Tris-HCl pH 6.8
0.05% (w/v) Bromophenolblue

Table 4.14: SDS PAGE running buffer

25mM Tris-HCl
200mM Glycine
0.1% (w/v) SDS

Table 4.15: Stacking gel (5%)

H_2O	1.46 ml
1M Tris-HCl pH 6.8	250 µl
40% acrylamide	250 µl
10% (w/v) SDS	20 µl
10% Ammonium persulfate	20 µl
TEMED	2 µl

Table 4.16: Separating gel (10%)

H_2O	1.92 ml
1.5M Tris-HCl pH 8.8	1 ml
40% acrylamide	1 ml
10% (w/v) SDS	40 µl
10% Ammonium persulfate	40 µl
TEMED	1.6 µl

Table 4.17: Separating gel (12%)

H_2O	1.72 ml
1.5M Tris-HCl pH 8.8	1 ml
40% acrylamide	1.2 ml
10% (w/v) SDS	40 µl
10% Ammonium persulfate	40 µl
TEMED	1.6 µl

4.2.15 Western blotting

After finishing SDS PAGE separation of proteins, gel was transferred to semi dry Western blot (WB) cell (BioRad). Filtration papers and nitrocellulose membrane (Whatman) were soaked in WB transfer buffer and placed on anode in following order: 3 pieces of filtration paper, nitrocellulose membrane, gel, 3 pieces of filtration paper, and covered with cathode. WB ran in WB transfer buffer at 10 V for 1 hour. Nitrocellulose membrane was then washed for 3 times with PBST

(1x conc. PBS supplemented with 0.05% (v/v) Tween) and blocked in 5% milk solution in PBST at 4 °C overnight. On the next day, membrane was washed with PBST again and incubated with primary antibody solution in 1% milk in PBST at room temperature for 1 hour. After additional series of washing with PBST, membrane was incubated with secondary antibody solution in 1% milk in PBST at room temperature for 30 minutes and washed with PBST again. Membrane was scanned using Odyssey Infrared Imaging System (LI-COR).

Table 4.18: Western blot transfer buffer

25mM Tris-HCl
200mM glycine
20% (v/v) methanol

4.2.16 Horizontal agarose gel electrophoresis

Nucleic acids samples (PCR products, products of restriction digestion and RNA isolation) were analyzed on 1% agarose gel (w/v). Agarose was dissolved in TBE buffer. Before polymerization, gel was heated and supplemented with GelStar (1:10 000 dilution, Cambrex). After loading of samples in 6x concentrated DNA Loading Dye (Fermentas), gel ran in 1x concentrated TBE buffer at 80 V for approximately 30 minutes to 1 hour. Gels were exposed and captured using Gel Doc XR+ (BioRad).

Table 4.19: 10x concentrated TBE buffer, diluted on 1x concentration with water

0.89M Tris base
0.89M boric acid
20mM EDTA

4.2.17 Urea RNA gel

RNA co-purified with IP samples was analyzed by urea RNA gel. At the beginning, 19.2 g of urea (final concentration 7M) was mixed with 4 ml of

10x concentrated TBE, 13.3 ml of 30% acrylamide and 5 ml of water. The mixture was then gently heated and stirred until urea was dissolved. Water was added to the final volume (40 ml), the mixture was cooled to room temperature and APS and TEMED were added to the gel. Gel was filled between the glasses and let to polymerize. After polymerization, gel was pre-run at 500 V for 15 minutes. Samples (10 µl of beads from IP preparation or 10 µl of input mixed with 6 µl of sample buffer) were loaded and gel ran at 500 V for 3 hours. Fixation of the gel was performed afterwards (40% methanol, 10% acetic acid twice for 30 minutes). Gel was then treated with dichroman for 10 minutes (3.4mM $K_2Cr_2O_7$ + 3.2mM HNO_3) and washed with water for 3 times, each wash for 30 seconds. Silver staining was performed (12mM $AgNO_3$ for 30 minutes) and after washing with water again, gel was developed (280mM Na_2CO_3 + 0.02% formaldehyde). Development was stopped upon addition of 1% acetic acid.

Table 4.20: RNA gel (10% acrylamide, 7M urea)

urea	19.2 g
10x TBE	4 ml
30% acrylamide	13.3 ml
10% Ammonium persulfate	400 µl
TEMED	25 µl
Total volume	40 ml

Table 4.21: Sample buffer for RNA gel

20mM Tris pH 8.0
8M UREA
Xylen Blue

4.2.19 Fluorescence resonance energy transfer (FRET)

Cells for FRET microscopy were grown in 6 well plates on cover slips and transfected with combinations of 3 constructs (see Results section, chapter 5). Total amount of transfected DNA was 2 µg per well (1.33 µg of RFP, 0.33 µg of CFP and 0.33 µg of YFP construct). After 18 - 24 hours from transfection, cells were fixed by 4% paraformaldehyde/PIPES for 10 minutes. After washing with PBS and water, cover slips were mounted by glycerol containing 2.5% 1.4-diazabicyclo[2.2.2]octane (DABCO). FRET microscopy was performed on Leica TCS SP5 confocal microscope. HCX PL APO 63.3/1.40-0.6 oil, Lbd Blue CS

objective, 50-mW diode laser (405 nm) and 100-mW Ar laser (514 nm) were used for all of the experiments. Data were acquired using 512 x 512 format and 200 Hz scan speed in 16 bit resolution. CFP was detected by 405 nm laser in wavelength range 446–506 nm and YFP by 514 nm laser in wavelength range 526–568 nm. Both lasers were set at 5 % of maximum intensity. The gain of photomultiplier detectors was adjusted to obtain appropriate dynamic range, but always at the same value for both lasers (800–950 V). FRET was measured using Acceptor Photobleaching Wizard (Leica). After donor and acceptor definition, acceptor (YFP) was bleached in the region of nucleoplasm by 514 nm laser on 60 % intensity by 3 to 4 scans. CFP intensity was measured before and after the bleach and FRET efficiency was calculated according to following equation:

FRET efficiency (%) = $(CFP^{after} - CFP^{before})x100/CFP^{after}$

15 to 25 cells were measured per each combination of vectors and averages were calculated. Three independent biological experiments were performed and after calculation of average of these repetitions, standard error of the mean was determined for each dataset. To decide on significance of obtained results, t-tests were performed, comparing two datasets (sets of averages of independent biological experiments for each dataset): mutant versus WT datasets.

5. Results

5.1 Splicing of β-globin pre-mRNA

To analyze dynamics of intron recognition we first decided to employ Fluorescence recovery after photobleaching (FRAP) and measure kinetics of different splicing factors at a splicing reporter gene. The splicing reporter we used is called E3 and is based on the β-globin gene that is under control of viral CMV promoter and contains a cyan fluorescence protein (CFP) fused to SKL tripeptide and 18 MS2 stem loop repeats downstream of β-globin coding sequence (Darzacq et al., 2006). The MS2 stem loops are bound by MS2 phage coat protein fused to fluorescent tag *in vivo* to enable visualization of the expressed pre-mRNA molecule. The purpose of SKL tripeptide is to direct the expressed protein to peroxisomes and thus not to disturb the nuclear fluorescence signal. Because the highest concentration of pre-mRNA in the nucleus is at the site of transcription, E3 pre-mRNA can be easily detected at the transcription sites together with associated splicing factors [see (Huranova et al., 2010)]. The interaction of splicing factors can thus be analyzed by fluorescent microscopy.

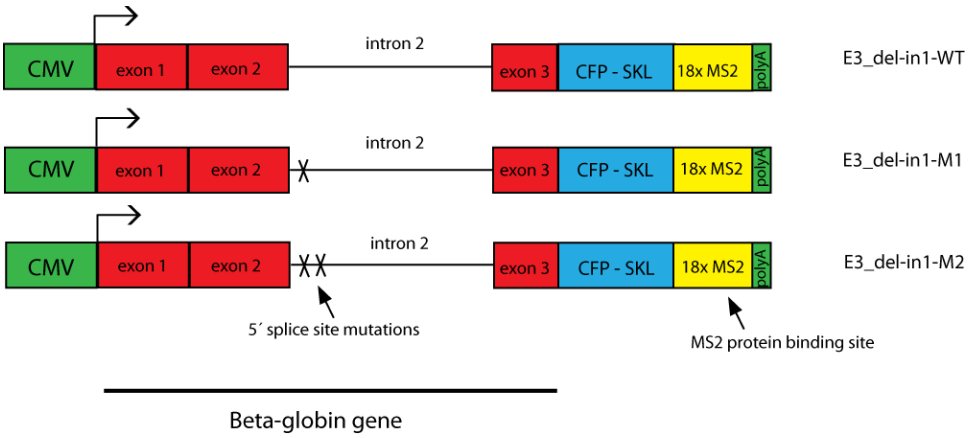


Figure 5.1: A set of vectors expressing β-globin gene. Vectors are under control of CMV promoters and, in addition to β-globin gene with one of the introns deleted, contain a sequence for CFP fused to SKL tripeptide (peroxisome import signal) and 18 MS2 stem loops (binding sites for MS2 phage coat protein). One of two series of vectors is depicted here (E3_del-in1), the other one (E3_del-in2) is symmetric.

To get to know more about influence of SS mutations on β-globin pre-mRNA splicing, a set of vectors, originated from E3 construct and expressing this gene was used (see figure 5.1 for E3_del-in1, analogous to E3_del-in2). The aim of this system was to establish stable cell lines in HeLa cells carrying an array of stably incorporated E3 vectors whose transcription would be visualized by MS2 coat protein fused to red fluorescence protein (RFP), encoded from a second vector.

The idea was to compare *in situ* dynamics of green fluorescence protein (GFP) tagged snRNPs on array of stably incorporated constructs expressing either wild type (WT) or mutated E3 pre-mRNA. Spliceosome assembly was expected to be altered in the presence of mutation of 5ss of intron 1 or intron 2. The eventually observed change in snRNPs dynamics determined by FRAP would allow us to judge on stepwise vs. penta-snRNP spliceosome assembly *in situ*. One or two mutations were introduced into 5ss as depicted at figure 5.2.

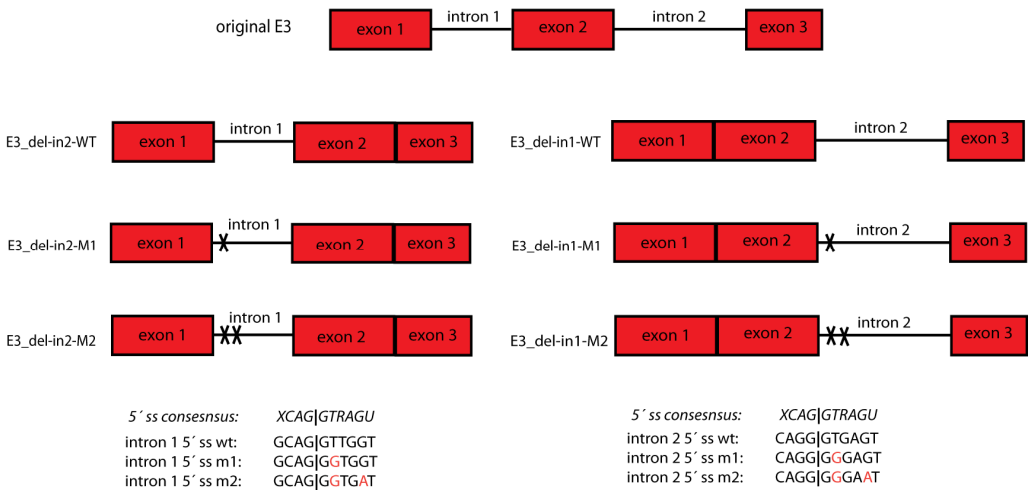


Figure 5.2: A scheme of manipulations which were performed with the E3 vector. After deletion of intron 1 or 2, two types of mutations were introduced in the 5ss of intron 2 or intron 1, respectively. The exact sequence, compared to mammalian 5ss consensus, is at the bottom, mutated nucleotides are in red. The exon-intron border is marked by vertical line.

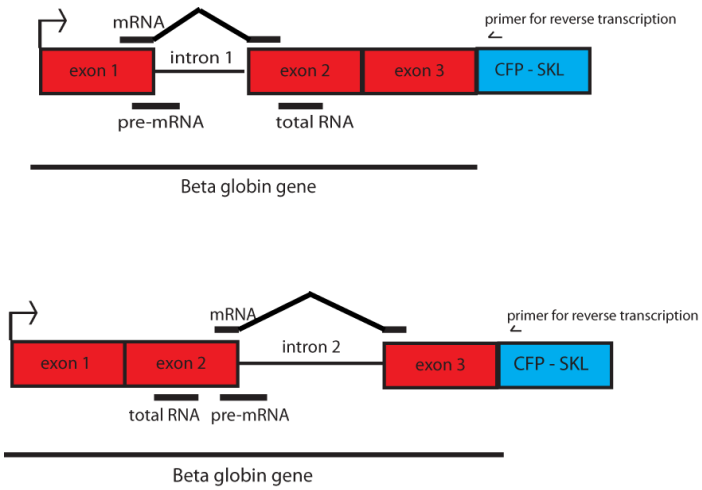


Figure 5.3: Sets of primers used for RT-PCR and qPCR analysis of splicing efficiency of E3 derived constructs. The horizontal lines represent location of the products which are amplified by the sets of primers. Primers for detection of total RNA are identical for both E3_del-in1 and E3_del-in2. All of the products have a length around 100 bp. In case of E3_del-in2 series (above), the intron retention results in formation of additional longer product (~230 bp) amplified by mRNA primers (see figure 5.4). The primer used for reverse transcription reaction is indicated as an arrow.

For each of the vectors, two types of mutations of 5ss were introduced. This resulted in a set of vectors referred as E3_del-in1-WT, E3_del-in1-M1, E3_del-in1-M2, E3_del-in2-WT, E3_del-in2-M1 and E3_del-in2-M2 (figure 5.2). In the first step, the vectors were transiently transfected into HeLa cell line and splicing was analyzed by RT-PCR and qPCR (figures 5.4, 5.5 and 5.6). Three sets of primers were used to detect the splicing products: for total RNA, pre-mRNA and mRNA (figure 5.3).

The shorter the intron is, the more sensitive it is to mutations introduced to the 5ss and the splicing efficiency is immediately decreased. Intron 1 (130 bp) appears to be less efficiently spliced in general, as the pre-mRNA product was detected with the E3_del-in2-WT vector. The longer intron 2 (1034 bp) responds to introduced mutations by cryptic SS activation instead (figure 5.4).

qPCR analysis revealed higher relative abundance of E3_del-in1-M2 mRNA than in case of E3_del-in1-M1. This reflects increased activation of cryptic SS with higher number of introduced mismatches into 5ss. The pre-mRNA product was only detected for E3_del-in1-M2 in RT-PCR which is in agreement with qPCR results (figures 5.4 and 5.5).

In case of E3_del-in2, the situation is slightly more complicated. The relative abundance of pre-mRNA exhibits the same pattern as observed in RT-PCR analysis, which means that the amount of unspliced form is increasing with the number of mutated nucleotides at the 5ss. The mRNA abundance appears to be the same for all of the vectors, regardless the introduced mutations, which is in conflict with the RT-PCR result (figures 5.4 and 5.6).

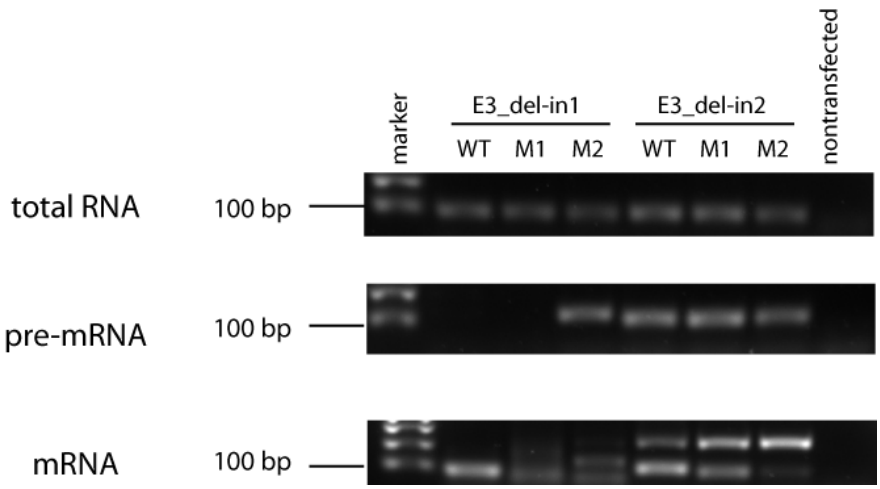


Figure 5.4: RT-PCR analysis of splicing efficiency of series of E3 derived vectors (agarose gel). All of the primers (figure 5.3) are designed for the size of products around 100 base pairs (bp). Interesting is the abundance of pre-mRNA product in E3_del-in2-WT column. In the mRNA line, possible cryptic SS activation is visible in columns of E3_del-in1 mutants with mRNA primers. In E2_del-in2 columns, the increase of unspliced form, detected by mRNA primers, is evident. The pre-mRNA primers detected equal amount of unspliced form for all E2_del-in2 vectors.

A function of E3 derived vectors in terms of microscopic visualization was examined by transient transfection to HeLa cell line stably expressing MS2 coat protein tagged with RFP. The presence of 18 high affinity binding sites on each

expressed mRNA molecule and high expression level of plasmid RNA due to strong CMV promoter resulted in majority of MS2-RFP molecules binding the transcripts. Those were than efficiently processed and exported to the cytoplasm. A picture of this is what can be seen on figure 5.7, showing that the nuclei of E3 transfected cells are empty (do not contain any RFP molecules).

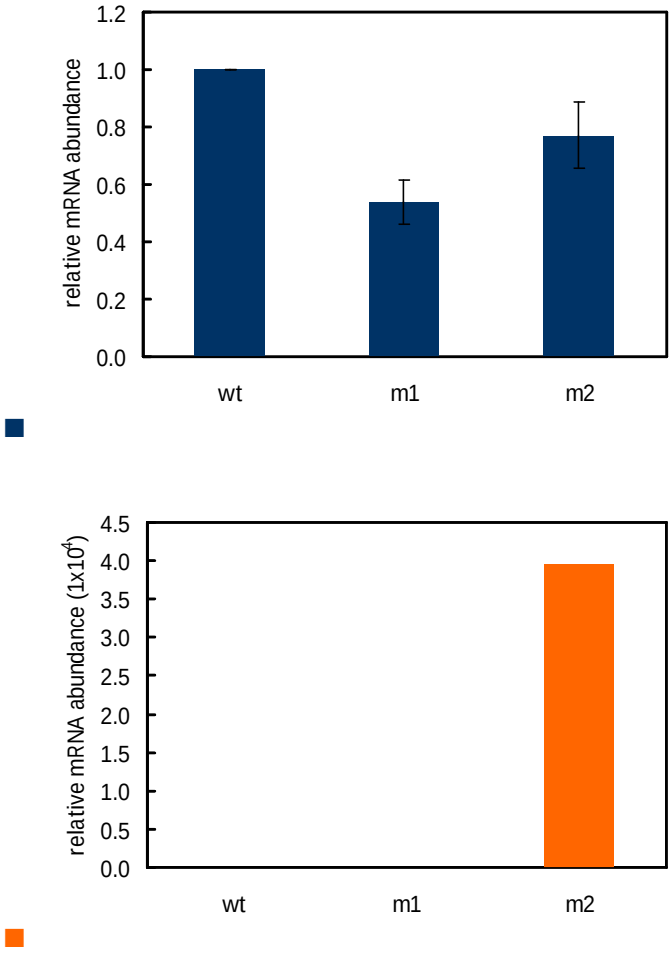


Figure 5.5: qPCR analysis of E3_del-in1 vectors splicing efficiency. The values (in case of mRNA abundance, an average of two experiments) represent the mRNA to total RNA (■) and pre-mRNA to total RNA ratio (■) which is normalized to ratio for WT construct.

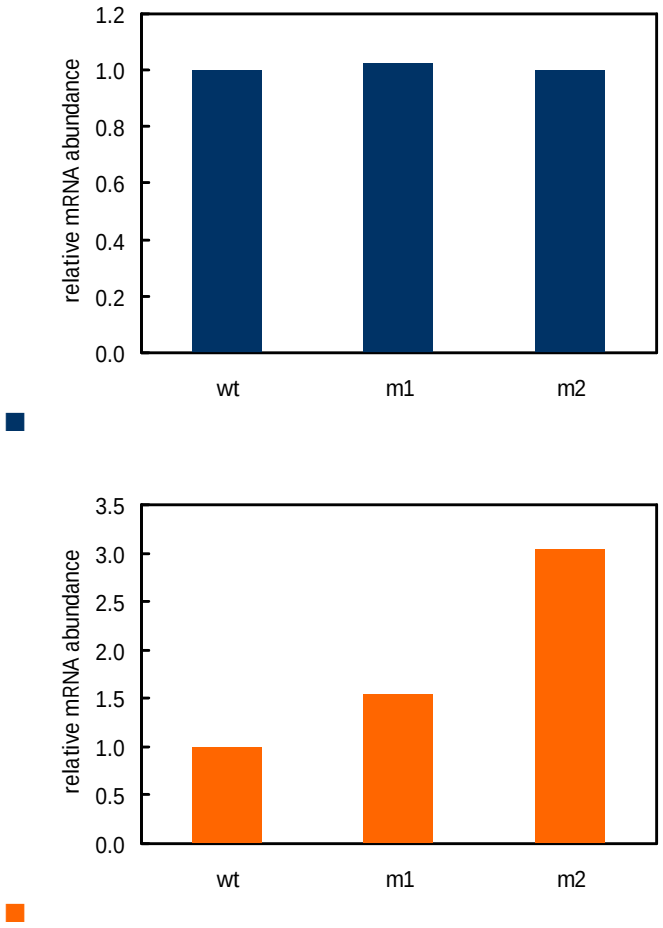
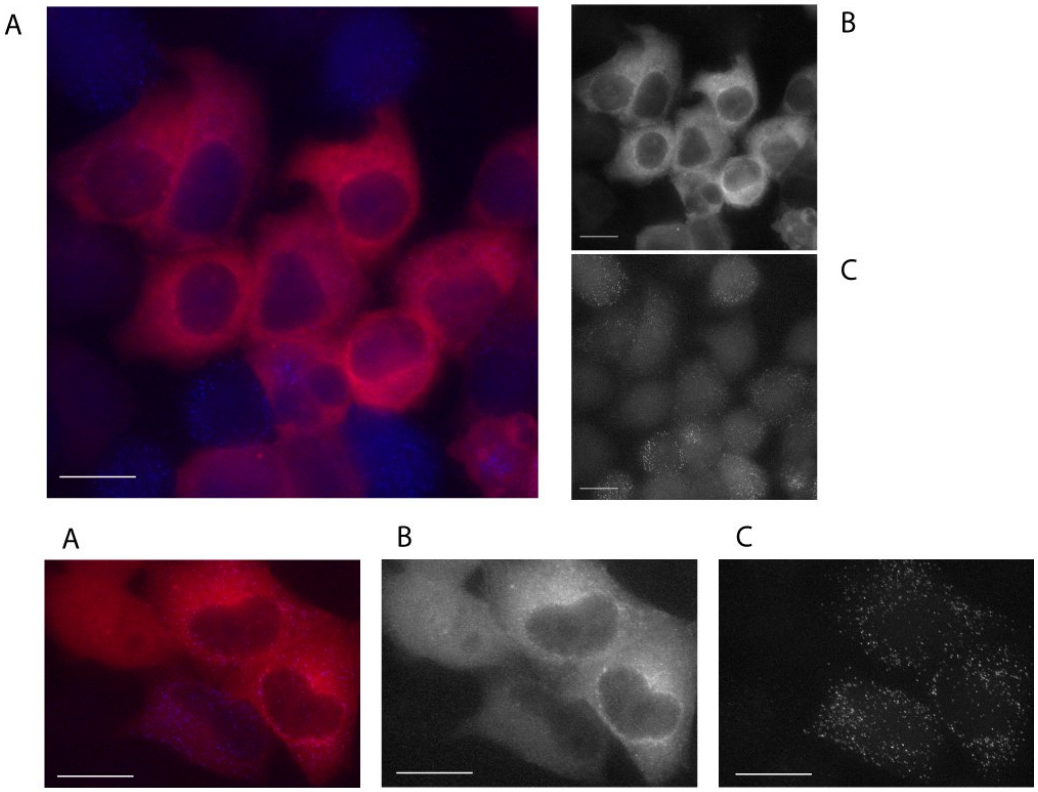


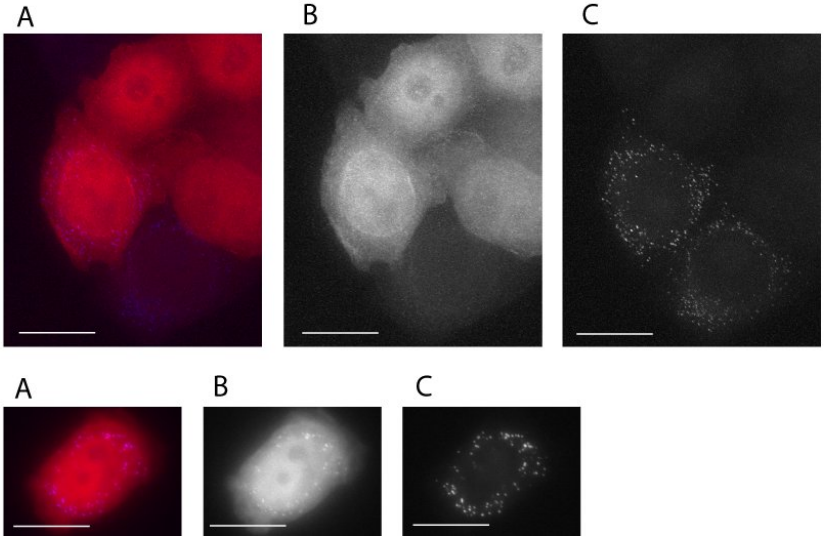
Figure 5.6: qPCR analysis of E3_del-in2 vectors splicing efficiency. The values (in case of mRNA abundance, an average of two experiments) represent the mRNA to total RNA (■) and pre-mRNA to total RNA ratio (■) which is normalized to ratio for WT construct.

On the next page: **Figure 5.7: Transient and stable transfection of E3 derived vectors to MS2-RFP stable cell line.** A – Merge. B – MS2-RFP, C – E3_del-in1-WT. HeLa cells stably expressing MS2-RFP vector were either transiently of stably transfected with E3_del-in1-WT, M1 and M2 vectors. The observed effect in transient transfection confirmed a proper function of MS2 stem loops placed downstream of β-globin gene. In stable cell lines, the dot in the nucleus presenting E3 transcription unit visualized by bound MS2-RFP is not visible. The establishment was thus not successful. Scale bar 10 μm.

transient transfection of MS2-RFP stable cell line



double stable cell line - MS2-RFP + E3_del-in1_WT



The next step was to establish the stable cell lines, expressing both MS2-RFP and E3 derived vectors. Since the E3 vectors do not carry any resistance gene, a co-transfection with plasmid encoding puromycin resistance gene was necessary. The selection and clonal expansion was followed by microscopy pattern evaluation. All of the expanded clones exhibited the right pattern of CFP, which localized to peroxisomes. Unfortunately, the transcription unit of incorporated E3 vectors marked by MS2-RFP was not visible in any of five analyzed clones (figure 5.7, bottom). Instead, the red signal was distributed all over the nucleus. This pattern probably reflected decreased or attenuated transcription of integrated E3 transgenes.

5.2 The E3 transcripts are bound by MS2 protein with high affinity

The unsuccessful stable cell lines establishment forced us to change strategy and employ alternative approaches. We took advantage of MS2 binding sites in our reporter and decided to purify protein-mRNA complexes using the MS2 system. First, we decided to examine the specificity of interaction between MS2 coat protein and the stem loops encoded by E3 vectors. Being strong enough, this interaction might allow us to isolate the complexes formed on pre-mRNA by biochemical methods and at least, to decide on snRNPs enrichment changes involved by 5ss mutations. The ability to purify the particular pre-mRNA would be the key to eventually perform such experiments.

We chose RNA immunoprecipitation (RIP) to examine the interaction. HeLa cells were transiently co-transfected with E3_del-in1-WT vector and a construct encoding MS2 coat protein fused to yellow fluorescent protein (YFP; MS2-YFP). Immunoprecipitation (IP) was performed with antibody against GFP and co-immunoprecipitated RNA was analyzed by qPCR with total RNA set of primers (see figure 4.3). The final number (values are normalized on input values) refers to fold enrichment over background (E3_del-in1-WT only transfected cells). The high amount of E3 RNA bound on the MS2 protein gave us information about the strong and specific interaction between the two elements. Based on this knowledge, we decided to further analyze the complexes precipitated with MS2-YFP using anti-GFP antibody.

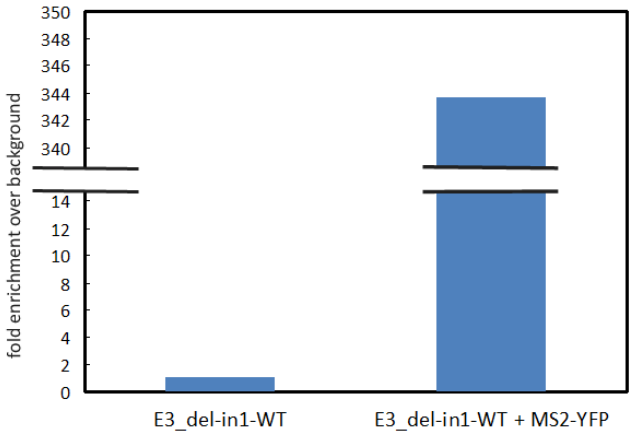


Figure 5.8: MS2 protein interaction with RNA molecule containing MS2 stem loops examined by RIP. IP with anti-GFP antibody was performed on cells transfected with either both MS2-YFP and E3_del-in1-WT or with E3_del-in1-WT only (background level) and co-precipitated RNA was analyzed by qPCR. The 350 times higher enrichment of E3 RNA on MS2 coat protein over background shows the high specificity of interaction between the RNA stem loops and the protein.

HeLa cells were transfected with MS2-YFP only and MS2-YFP plus E3_del-in1-WT. Non-transfected cells were used as a negative control. IP was performed with anti-GFP antibody (cross-reactive also with YFP). First of all, we performed a Western blot (WB) with anti-GFP antibody to check if IP was working. In addition, we used Y12 antibody (detecting Sm ring of proteins, a core part of most of the snRNPs) to check whether snRNPs co-immunoprecipitated with our complex. The WB shown in figure 5.9 is the evidence for properly working IP (on the right). Nevertheless, the Y12 signal does not differ in non-transfected and MS2-YFP only transfected cells vs. cells transfected with both MS2-YFP and E3_del-in1-WT. This reflects the character of Sm protein complexes which easily attach to sepharose beads used in IP.

We obtained similar results when performing RNA gel (figure 5.10) with the same set of samples. The snRNA molecules as well as other detectable types or RNA were detected at a similar level in all of the samples (MS2-YFP transfected as well as non-transfected cells). This highly sensitive method led us to a conclusion

that we were not able to isolate snRNPs containing complexes specifically bound to β-globin mRNA molecule using MS2 coat protein co-IP from entire cellular environment. When analyzing directly single proteins of snRNPs by WB, particularly U2B´´, a protein component of U2 snRNP, we were not able to detect any signal in any sample (data not shown).

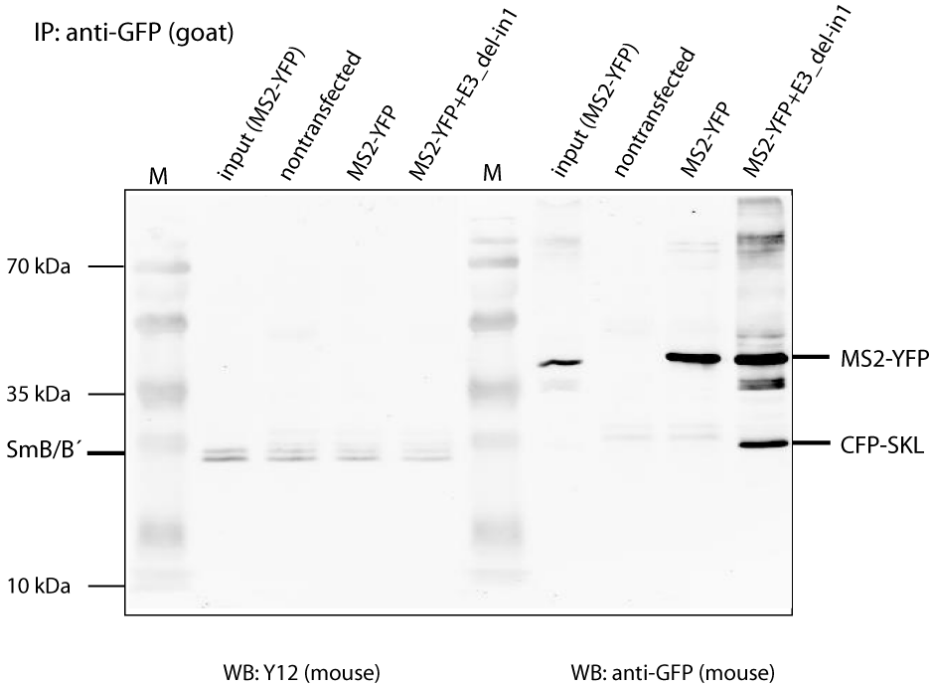


Figure 5.9: Western blot analysis of co-immunoprecipitated complexes of snRNPs bound on E3 transcript (left). IP check (right). The IP was working properly. Bands around 45 kDa respond to MS2-YFP fusion proteins while lower band appearing specifically in E3_del-in1-WT transfected cells is a protein product of CFP-SKL expressed from E3 vector. On the left, the signal of Y12 antibody does not differ among the columns and is even present in non-transfected cells.

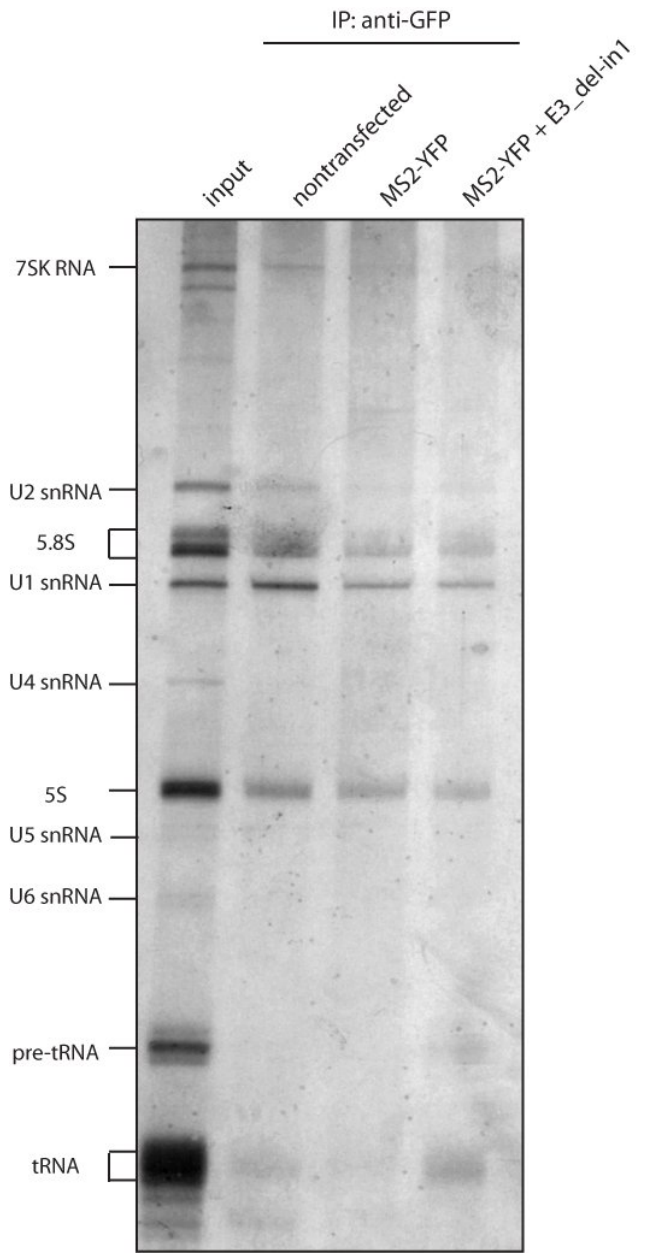


Figure 5.10: The RNA gel analysis of co-IP complexes. The RNA fraction of IP was analyzed by urea RNA gel. The abundance of snRNA molecules appears to be highest in non-transfected cells. U1 is the only one detected in all of the samples, including non-transfected cells. The detectable signal is thus unspecific.

5.3 Differential splice sites occupation detected by FRET

The next step in describing the early events of spliceosome assembly *in situ* was to employ a FRET approach. The architecture of spliceosomal complexes built on pre-mRNA molecule can be studied by this approach. Our collaborator Massimo Caputi from University of Florida provided us with a different set of constructs designed for FRET detection of proteins interacting with 5ss and 3ss. The vectors (WT ones will be referred as pRed-bGlo_5M and pRed-bGlo_3M, see figure 5.11) are based on intron 1 of β-globin gene and border sequences of adjacent exons (exons 1 and 2, for the exact sequence inserted into the constructs see Materials and methods section). In addition to this, the vectors contain hcRED fluorescent protein coding sequence (upstream from the β-globin gene segment) and a single MS2 stem loop which is localized close to either 5ss (pRed-bGlo_5M) or 3ss (pRed-bGlo_3M) of β-globin intron 1 to allow a detection of MS2 protein and SS bound protein interaction (see figure 5.11 for vectors scheme). Because energy transfer during FRET occurs at short (≤ 10nm) distances only, the inserted exon sequences need to be short to keep a small distance between MS2 site and SS. Proteins tagged with appropriate fluochrochromes and simultaneously bound to MS2 and SS are then capable of FRET reflecting a specific interaction of particles which are in close proximity. Based on previous reports, a single MS2 binding site is sufficient for FRET detection (Huranova et al., 2009). We used acceptor photobleaching FRET method to detect the interactions within the protein pair (see Materials and methods).

The proteins which are known to specifically bind 5ss (U1C) and 3ss (U2AF35) sequences were used and tagged with CFP. A second member of the pair was MS2-YFP. Either GU or AG dinucleotide of β-globin intron 1 was substituted for either AA or TT, thus aborting one of the SS sequences. These manipulations resulted in following set of constructs: pRed-bGlo_5M (WT SS), pRed-bGlo_3m5M (3ss mutation) and pRed-bGlo_5m5M (5ss mutation) for detection of interactions at the 3ss and pRed-bGlo_3M (WT SS), pRed-bGlo_5m3M (5ss mutation) and pRed-bGlo_3m3M (3ss mutation) for detection of interactions at the 5ss (see figure 5.12).

Triple transfection was performed including both proteins of the FRET pair as well as the RNA target (figure 5.13). Significantly lower expression levels of pRed-bGlo_3M or 5M constructs in comparison to empty pRed vectors indicated

impaired or disadvantaged processing of hcRED protein when coupled by β-globin gene segment. Indeed, when RT-PCR analysis was performed, the constructs exhibited inefficient splicing regardless of WT or mutated SS (figure 5.14).

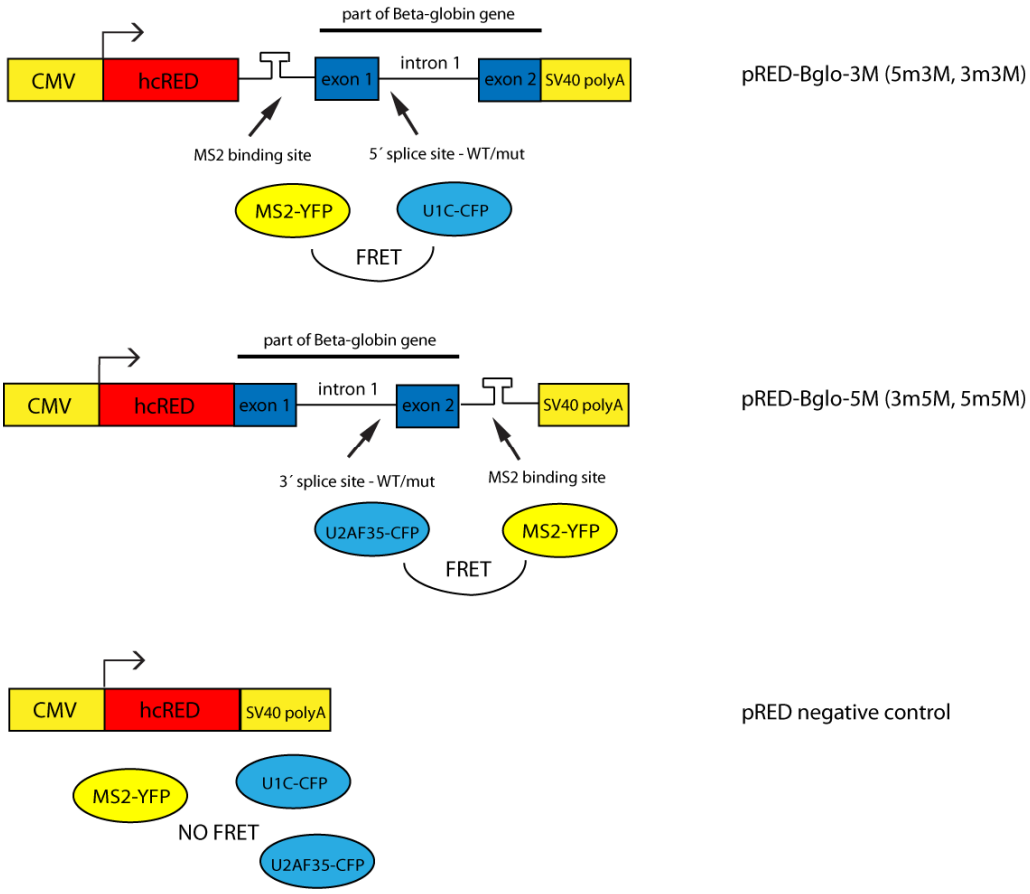


Figure 5.11: A scheme of constructs used for FRET. Each of the vectors contains a fragment of β-globin gene, a single MS2 stem loop and hcRED protein coding sequence. The 3M series was used for 5ss interactions (U1C-CFP + MS2-YFP) and 5M series for 3ss interactions (U2AF35-CFP + MS2-YFP) detection. A vector lacking any β-globin sequence or MS2 binding sites was co-transfected with FRET pairs as a negative control.

The FRET experiments were carried out in three independent biological repetitions, each containing 15 to 25 cells. In the first set of measurements, the positive controls (3M, 5M) and negative control (empty pRed vector) were transfected. Obtained FRET values for U1C-CFP and U2AF35-CFP plus MS2-YFP were evaluated to get an idea about possible background levels. Microscope settings were established to successfully distinguish false positive measurements. FRET was measured in nucleoplasm of the cells, carefully selecting cells with equal expression levels of CFP and YFP fluorophores.

In the second round, FRET measurements of protein pairs on mutated RNA scaffold were performed. After confirming that 5ss and 3ss mutations significantly decrease FRET values for U1C and U2AF35 (5m3M and 3m5M constructs), the opposite SS mutations for each vector were established (resulting in 3m3M and 5m5M constructs, see figure 5.12). Remarkably, a decrease in FRET values was observed for both U1C and U2AF35, indicating a possible role for opposite SS in E complex components recruitment to their binding sites on pre-mRNA. For U2AF35, statistically significant ($p = 3.8$ % for 3ss mutation vs. WT, $p = 4.7$ % for 5ss mutation vs. WT) effect was observed when 3 biological experiments were averaged (figure 5.16). A slightly more complicated situation was observed for U1C: while within two single biological experiments, data suggested an effect of 3ss mutation on U1C recruitment (although different absolute values were observed for each of the repetitions), the effect was not observed by the third performance. The final effect is thus obvious but not statistically significant ($p = 26.1$ % for 5ss vs. WT, $p = 23.28$ % for 3ss vs. WT) (figure 5.17).

Illustrating records of FRET for each pair and each pre-mRNA scaffold are on figures 4.18 and 4.19. A typical cell nucleus characterizing the particular trend line was chosen from each dataset.

5.4 RIP analysis of splice sites attached complexes

The interesting FRET data, especially the significance of finding that U2AF35 recruitment to 3ss AG dinucleotide requires functional 5ss on the opposite site of the intron, as detected *in vivo*, led us to assumptions on the situation within the second β-globin intron. Since we did not possess constructs containing intron 2

suitable for FRET measurements, we decided to employ RIP experiments to check the pathways of snRNPs recruitment to intron 2 SS.

We transiently co-transfected U2B´´-GFP construct with E3_del-in1-WT and E3_del-in1-M2 constructs. Single transfection of either E3_del-in1-WT or U2B´´-GFP was used as a negative control. IP with antibody against GFP was performed and equal amount of precipitated U2B´´-GFP for both WT and mutated E3 vectors co-transfection was checked by WB (figure 5.20). Pull down efficiency was similar for all of the samples.

A band around 55 kDa represents U2B´´-GFP fusion protein. The additional lower bands detect CFP expressed from E3 derived vectors. Interestingly, the expressed proteins differ between WT and M2 vectors, since the 5ss mutation apparently activates cryptic splice sites which probably results in aberrant protein production. The 45 kDa product expressed from WT construct might be β-globin-CFP-SKL fusion protein while cells transfected with the mutated one lack this band and only produce CFP-SKL protein (25 kDa) (figure 5.20).

Co-immunoprecipitated RNA was analysed by qPCR using total RNA set of primers (figure 5.3). The final numbers were normalized on input values. Enrichment over background level (E3_del-in1-WT only transfected cells) is plotted. RIP with U2B´´-GFP and E3 derived vectors was performed twice and average of these two experiments was calculated. High standard errors of the mean reflect the character of the method employing transient transfections etc. Nevertheless, we managed to detect a difference in U2B´´ occupancy on intron with WT 5ss compared to mutated 5ss carrying intron. This is in agreement with FRET results and brings further evidence of requirement of 5ss integrity for U2AF and U2 snRNP recruitment to 3ss (figure 5.21).

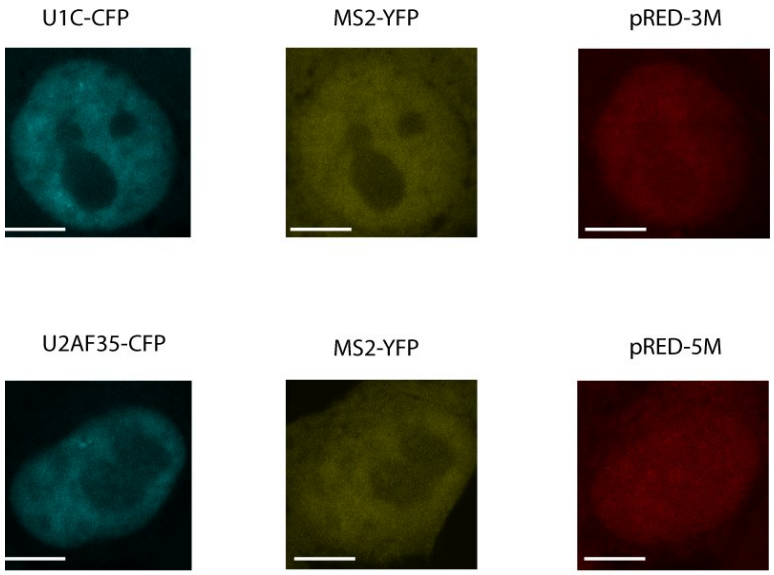


Figure 5.12: An example of triple transfections for both FRET pairs. Scale bar 7.5 μm.

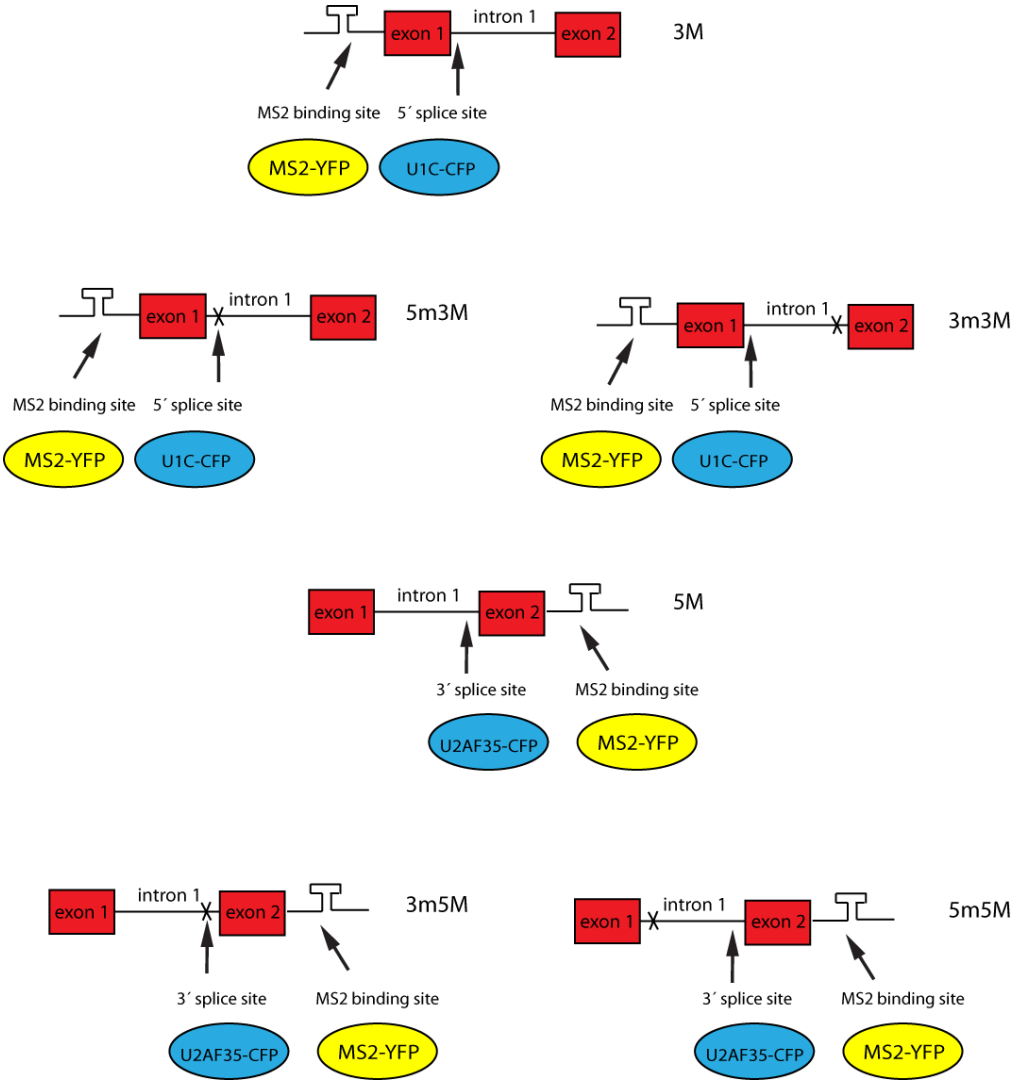


Figure 5.13: A complete scheme of constructs used as scaffolds for FRET interactions. Shortened names are used here for easier orientation. 3M series was used for 5ss interactions detection (U1C-CFP + MS2-YFP) and 5M series for 3ss interactions detection (U2AF35-CFP + MS2-YFP). Arrows indicate the binding sites of protein factors on the RNA scaffold. Crosses indicate sites of mutations introduced by site directed mutagenesis.

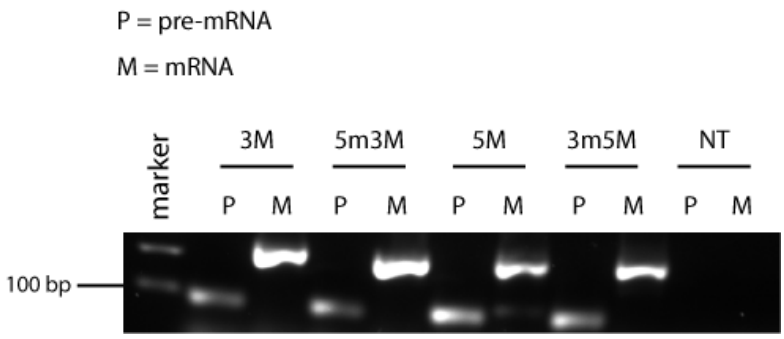


Figure 5.14: RT-PCR analysis of FRET scaffold constructs splicing (agarose gel). These constructs contain intron 1 and short fragments of adjacent exons of β-globin gene. Pre-mRNA (P) and mRNA (M) primer pairs were used to detect the products of splicing. The mRNA products are supposed to be 100 bp in length but appearance of the longer product (~230 bp) points to the intron retention. (A slight band in 5M M column is due to a pipeting mistake).

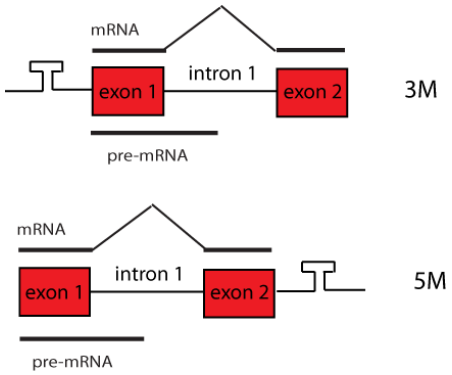


Figure 5.15: Sets of primers used for RT-PCR analysis of splicing of 3M and 5M constructs. The horizontal lines represent the amplified products, the desired ones of length around 100 bp. When intron is retained the length of products amplified by mRNA primers increases to ~230 bp.

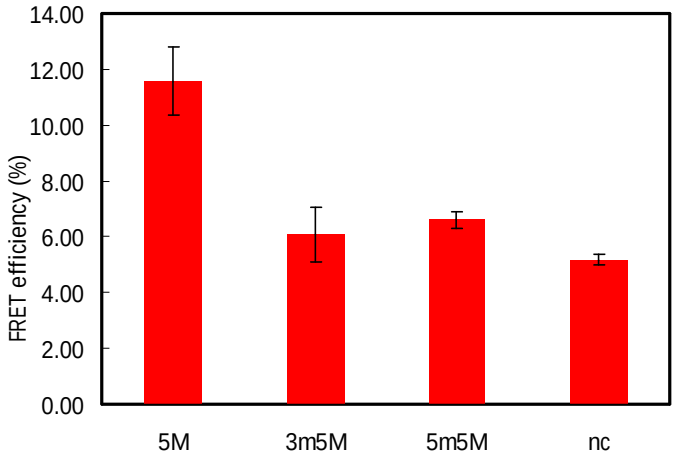


Figure 5.16: Detection of U2AF35 interaction with 3ss by FRET. FRET efficiency (average of 3 biological experiments) for U2AF35-CFP and MS2-YFP fluorescent pair. Error bars represent standard error of the mean. 5M (a WT RNA scaffold) serves as a positive control. 3m5M = mutated 3ss, 5m5M = mutated 5ss. pRed empty vector is for negative control (nc).

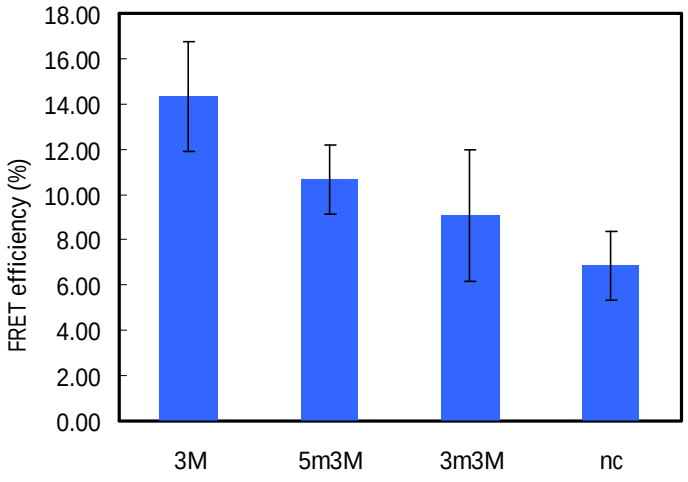
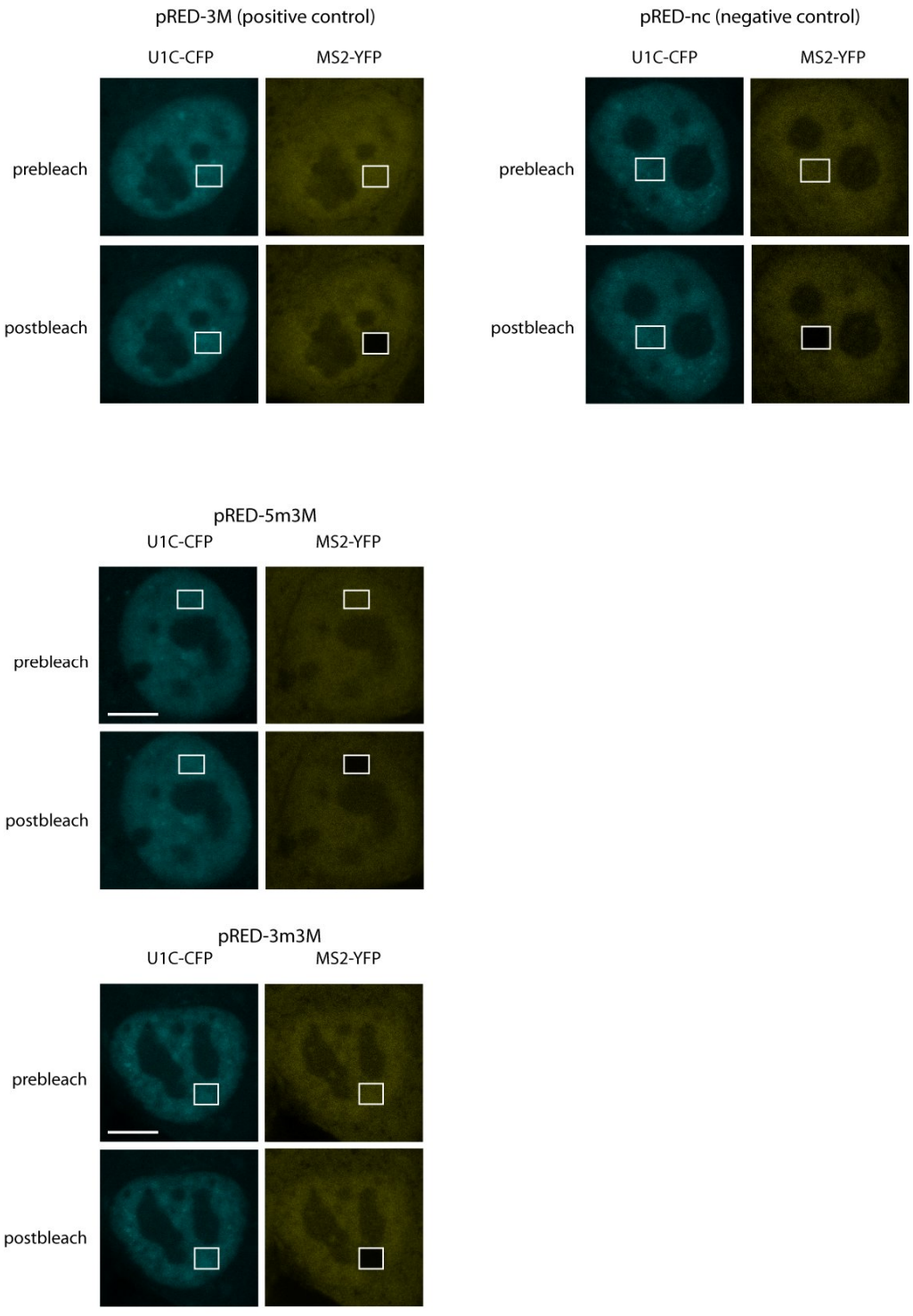
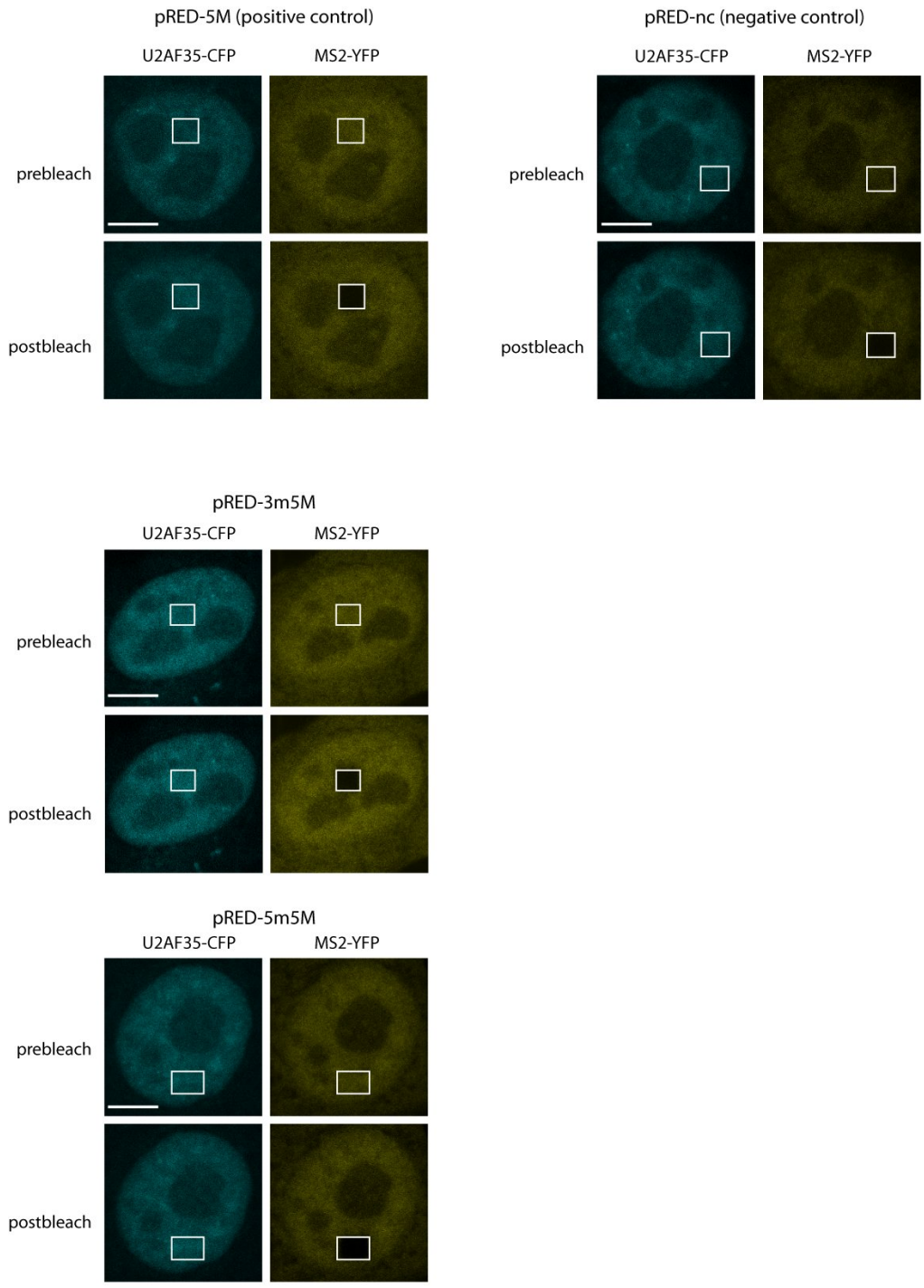


Figure 5.17: **Detection of U1C interaction with 5ss by FRET**. FRET efficiency (average of 3 biological experiments) for U1C-CFP and MS2-YFP fluorescent pair. Error bars represent standard error of the mean. 3M (a WT RNA scaffold) serves as a positive control. 5m3M = mutated 5ss, 3m3M = mutated 3ss. pRed empty vector is for negative control (nc).


pRED-3M (positive control)
U1C-CFP
MS2-YFP
prebleach
postbleach
pRED-nc (negative control)
U1C-CFP
MS2-YFP
prebleach
postbleach
pRED-5m3M
U1C-CFP
MS2-YFP
prebleach
postbleach
pRED-3m3M
U1C-CFP
MS2-YFP
prebleach
postbleach

On previous page: **Figure 5.18: Representative FRET records for U1C-CFP and MS2-YFP pair.** Acceptor photobleaching method was used to obtain these results. Increase in donor (CFP) fluorescence is detectable when acceptor (YFP) is bleached in the positive control. Note the low level of donor (CFP) fluorescence increase after acceptor (YFP) bleach for both SS mutants (5m3M and 3m3M). FRET values are at the level of negative control for both mutants. The chosen cells are within the trend line of the dataset, thus demonstrating a comparison of examined mutants with positive and negative control. It is notable that average values of mutants differ among single biological repetitions in case of U1C. These records therefore represent a trend of the repetition in which the effect was observed. Scale bar 7.5 μm.

On the next page: **Figure 5.19: Representative FRET records for U2AF35-CFP and MS2-YFP pair.** Acceptor photobleaching method was used to obtain these results. Increase in donor (CFP) fluorescence is detectable when acceptor (YFP) is bleached in the positive control. Note the low level of donor (CFP) fluorescence increase after acceptor (YFP) bleach for both SS mutants (3m5M and 5m5M). FRET values are at the level of negative control for both mutants. The chosen cells are within the trend line of the dataset, thus demonstrating a comparison of examined mutants with positive and negative control. Scale bar 7.5 μm.


pRED-5M (positive control)
U2AF35-CFP
MS2-YFP
prebleach
postbleach
pRED-nc (negative control)
U2AF35-CFP
MS2-YFP
prebleach
postbleach
pRED-3m5M
U2AF35-CFP
MS2-YFP
prebleach
postbleach
pRED-5m5M
U2AF35-CFP
MS2-YFP
prebleach
postbleach

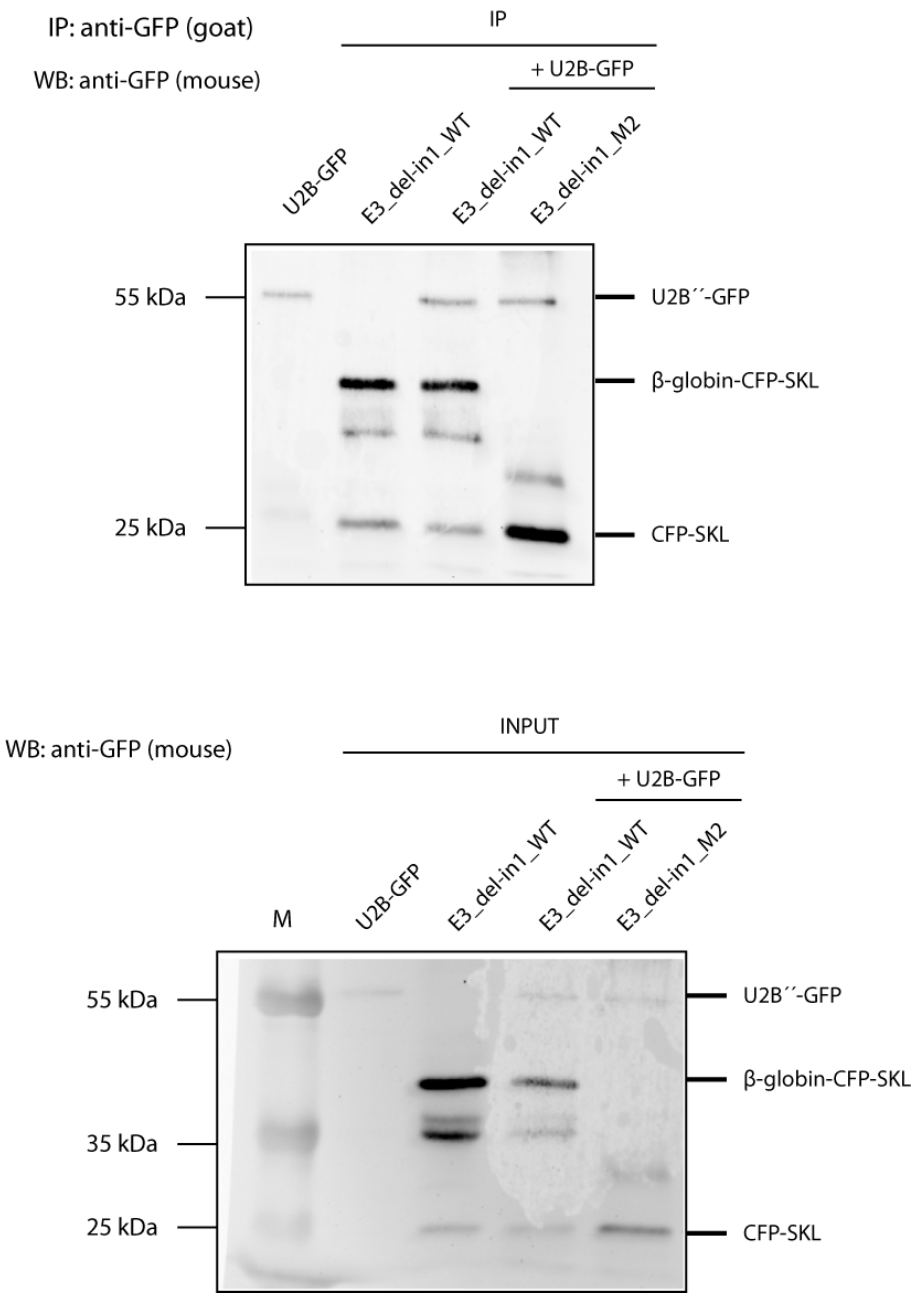


Figure 5.20: Pull down control for RIP experiment analyzing U2B´´-GFP enrichment on RNA expressed from E3 derived vectors. IP – upper panel. Input - lower panel. The immunoprecipitated proteins were checked on WB with anti-GFP antibody. U2B´´-GFP pull-downed protein is on 55 kDa. Around 40 kDa, a putative β-globin-CFP protein is visible (only expressed from WT E3 derived constructs). 25 kDa band represents CFP-SKL protein expressed from all types of E3 vectors.

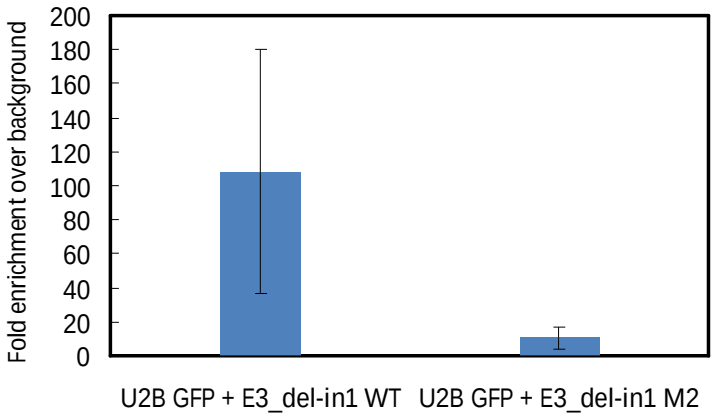


Figure 5.21: U2B´´ enrichment on β-globin intron 2 with WT vs. mutated 5ss analyzed by RIP (average of two biological repetitions). Error bars represent standard errors of the mean. Cells were co-transfected with U2B´´-GFP and either E3_del-in1-WT or -M2. IP was performed with anti-GFP antibody. Transfection of E3_del-in1-WT only was used to determine a level of background. qPCR obtained information about of co-precipitated RNA was normalized on input and on background level.

6. Discussion

We have investigated splicing efficiency of β-globin introns within E3 splicing reporters. We found that shorter intron 1 is more susceptible to remain unspliced. A fraction of unspliced form was detected when the intron was surrounded by full-length exons and both SS were retained. Splicing was completely abolished when short pieces of adjacent exons accompanied intron 1. This is in agreement with the findings of multiple enhancer sequences localized in exon 1 and exon 2 of β-globin gene. These elements are responsible for SR proteins SC35 and ASF/SF2 recruitment to the pre-mRNA molecule and enhancement of the splicing efficiency (Schaal and Maniatis, 1999). SR proteins generally play an important role in increasing the efficiency of splicing and SS recognition. The dependence of β-globin intron 1 splicing on ESE recruited SR proteins evident from our RT-PCR analysis can possibly be examined in the future by tethering of SR proteins fused to MS2 coat protein to our FRET constructs which benefit from MS2 stem loops placed adjacent to 5ss or 3ss.

Intron 2 has more than 1000 bp in length and appears to be more efficiently spliced. The introduced mutations activate cryptic 5ss within the intron, which influences the protein products and has unknown impact on spliceosome assembly. The 5ss mutations have an effect on construct expression in general. The expression is decreased when the mutations are introduced to the 5ss (data not shown), which confirms the concept of 5ss impact for transcription stimulation and U1 snRNA recruitment to the transcription unit to possibly couple transcription and splicing processes [(Spiluttini et al., 2010) and references therein].

Splicing of multiple introns within single gene is proposed to occur in 5′ to 3′ direction which reflects the co-transcriptional nature of splicing. However, some exceptions have been identified suggesting that certain downstream introns can exhibit higher efficiency of co-transcriptional splicing, regardless the intron strength (Pandya-Jones and Black, 2009). Our results reveal differences between the two constitutively spliced introns of β-globin pre-mRNA. The first intron appears to be less efficiently spliced. Nevertheless, the situation should be checked when both introns are present in the transcript. With usage of appropriate primer for reverse transcription placed downstream of polyadenylation signal, the native pre-mRNAs

(still attached to RNA polymerase II) should be analyzed to decide on which of the introns is the first to be spliced, if any of the two.

The unsuccessfulness of E3 stable cell lines establishment led us to alternative approaches. There might be several reasons for inefficient stable cell lines production. The construct might have been integrated into heterochromatin regions which could have led to partial or full silencing of the transgene. This could have been active for a restricted period during which the CFP-SKL was expressed and transported into peroxisomes where it was detected. CFP-SKL might have had long turnover period within peroxisomes and thus allow detection of the protein product by the time the transgene was not active any more. Since the CFP signal was only detected in a small number of cells within puromycin resistant population, it can by assumed that the resistance encoding vector was, under the selection pressure, preferentially integrated into HeLa cells genome while the E3 vectors were turned off or even not integrated at all. For future attempts to establish stable cell line like this, the ratio of resistance carrying vector vs. vector carrying gene of interest should be carefully determined and empirically examined to be efficient. All of the transgenes should integrate into a single genome place and overabundance of total DNA which the cells were transfected with might have had a negative impact on introduction of foreign DNA to the genome. The resistance encoding plasmid is in addition preferential for the cells to be maintained since they are exposed to selection pressure. It might be useful to follow the original procedure used for E3 cell lines establishment (Darzacq et al., 2006), involving electroporation of transgene DNA instead of lipofection we used. Electroporation ensures that a limited number of constructs efficiently enters the cells and becomes stably integrated. Another point is the choice of constitutive CMV promoter. This may be too active and strong and can induce heterochromatin formation around foreign DNA which does not have any purpose to be expressed. Solution to this might be the usage of inducible promoter which is activated upon doxycycline (or some other drug) addition. Alternative approaches are to use a completely different delivery method, for example lentiviral vectors or FLP/FRT system employing a single target sequence for transgene integration and recombinase encoded from additional vector.

We have brought a new insight into early steps in spliceosome assembly *in vivo* by FRET. pRed-bGlo constructs that do not posses full length exons

are inefficiently spliced. However, we were able to detect U1 and U2AF interactions with pre-mRNA indicating that splicing is inhibited in later steps of spliceosome assembly. Inefficient splicing offers advantageous conditions in (likely) prolonged U1 and U2AF interactions with target RNA. We were thus able to detect standby but weak interactions of E complex on intron lacking adjacent exonic enhancer elements. The absence of enhancer elements possibly affects SR proteins interaction with the transcript.

SR proteins are known to participate in SS associated complexes as well as in cross exon interactions of snRNPs contributing to exon and subsequent intron definition (Fu and Maniatis, 1992; Staknis and Reed, 1994; Wu and Maniatis, 1993; Zuo and Maniatis, 1996). Our system is predicted to lack this level of complexity, as well as any cross exon interactions which are routinely utilized in mammalian transcripts. The existing cross intron interactions might be less stable and more sensitive to any disturbing conditions as a result of that. Since we do not detect any spliced mRNA originated from our constructs, we assume that only early spliceosomal complex, not capable of commitment to splicing, is assembled on the target RNA molecule.

Important is the finding that U2AF35 recruitment to 3ss is remarkably compromised in absence of intact 5ss. U2AF35 recognizing 3ss forms heterodimers with U2AF65 subunit recognizing PPT with cooperation of SF1 binding to BPS (Berglund et al., 1998). These elements (BPS, PPT, 3ss) are known to be localized close together in E complex (Kent and MacMillan, 2002) which implies also close contacts of involved proteins. In addition, blocking of 5′ end of U1 snRNA and U1 snRNP depletion from the splicing reaction remarkably compromises U2AF65 recruitment to the pre-mRNA (Cote et al., 1995; Li and Blencowe, 1999). These experiments bringing evidence for U2AF dependence on U1 snRNP associated with pre-mRNA were performed *in vitro*. Our results were obtained *in vivo* as the first evidence for this phenomenon to occur in living cells.

The existence of cross intron interactions is a condition necessary for efficient splicing reaction. These interactions involve a set of proteins recognizing 5ss (U1 snRNP), BPS (SF1), PPT (U2AF65) and 3ss (U2AF35). Such protein bridge oven intron is conserved in both yeast and mammals (Abovich and Rosbash, 1997). The entire intron network in mammals is connected by additional factors as FBP11

and members of SR proteins family (Kent et al., 2005) and can be more complicated than the yeast one. The absence of one factor can possibly lead to certain obstructions in the other factors recruitment. Since some of the E complex proteins are considered to recruit independently on U2AF, a compromised but not completely aborted interaction of U1C with 5ss in absence of 3ss observed in our FRET experiments might suggest that there is some redundancy in E complex formation regarding 3ss availability. The early interactions with 5ss require neither functional BPS nor presence of U2AF *in vitro* (Kent et al., 2005). There probably is some dynamic interplay among SS recognition factors *in vivo* which in turn results in slight alterations in U1 snRNP recruitment in absence of 3ss. What we detect by FRET (and what is determined by statistical character of FRET evaluation) might reflect the *in vivo* situation more precisely than nuclear extracts usage for spliceosomal complexes reconstruction.

SR family members ASF/AS2 and SC35 were both shown to interact with U1-70K and U2AF35 *in vivo*; HCC1 can form complexes with both subunits of U2AF (Ellis et al., 2008). Although it is not clear whether these interactions form across an exon or an intron and what are the sequence requirements for them to mediate the SS recognition, it is another evidence for U1 snRNP and U2AF communication *in vivo.* It also points to the connection of U2AF65 and 35 through another protein factor, admitting possibility of preformed complexes existence (Ellis et al., 2008). Functional networks of splicing factors which would be preformed *in vivo* might result in stronger requirements of intact pre-mRNAs to let each member of the complex efficiently bind to particular binding site. What we detect is architecture of spliceosomal complexes directly on pre-mRNA target. The exact course of splicing and the members of cross intron bridge built on our FRET constructs remain to be elucidated.

The strategy of splicing complex analysis by isolating pre-mRNA molecules containing MS2 stem loops was not functional for us. Despite routine usage of such approaches in *in vitro* systems for spliceosomal complexes purification, isolation of particular pre-mRNA molecule bound by snRNPs is not easy to perform when working with total cellular extract, although strong and specific interaction of MS2 coat protein with the target RNA was confirmed. The conditions kept for our IP experiments are fully operational for isolation of snRNP

components and analysis of the RNA content co-purified with them. This approach allows us to study early spliceosomal complexes assembled on our pre-mRNA in WT or mutated SS context.

β-globin pre-mRNA splicing requires ESE elements to be accomplished (Schaal and Maniatis, 1999). The RNA template we used for RIP experiments contains intron 2 accompanied by full-length exon 2 and a part of exon 3 and it undergoes efficient splicing. We assume that the splicing reaction on this pre-mRNA molecule is mediated by ESE dependent recruitment of SR proteins, in contrast to FRET vectors, and that E complex formation is followed by A complex which commits pre-mRNA to splicing. The interactions of snRNPs with the pre-mRNA molecule are thus strong enough for efficient isolation of RNA-protein complexes by RIP. On such pre-mRNA target, SR proteins can possibly contribute to U1 and U2 cooperation (Boukis et al., 2004).

We were able to detect intact 5ss requirement for U2 snRNP assembly on β-globin pre-mRNA by RIP. U1 snRNA – 5ss base-pairing requirement for U2 to bind pre-mRNA *in vivo* has been reported for yeast cells (Lacadie and Rosbash, 2005; Tardiff and Rosbash, 2006). U1 and U2 snRNPs were proposed to form a di-snRNP on introns of *Schizosaccharomyces pombe* and a contact between U1 and U2 was detected also in HeLa cells (Xu et al., 2004). The extension of cross intron interactions formed in E complex to the next stage of spliceosome assembly is likely to maintain the spatial arrangements of intron leading to splicing catalysis. Additionally, U2 was found to be associated with pre-mRNA already in the E complex and independently of BPS (Das et al., 2000) and might thus assemble into the existing network of SS interactions. U2 snRNA modified nucleotides were even shown to be necessary for E complex formation (Donmez et al., 2004) which further supports the idea of U2 function in the early stages of spliceosome assembly. Moreover, U2 is suggested to be involved in BPS independent but 5ss dependent interactions between U1 and U2AF65 to join the SS in E complex (Donmez et al., 2007). When U1 recruitment, the initial and key step in splicing, is compromised by 5ss mutation, the E complex formation is altered, as detected by FRET, and that probably results in decrease of U2 binding of the pre-mRNA and in a break of U1-U2 interactions across the intron. We thus confirmed SS cooperation at both E and A complex stages of spliceosome assembly *in vivo*.

In splicing of minor class of introns known as U12 (AT-AC) type, U11 recognizes the 5ss and U12 the BPS. Strikingly, these snRNPs were found to form a di-snRNP particle in HeLa cells (Frilander and Steitz, 1999). That suggests a parallel mechanism of SS recognition established in the evolution of splicing which might persist in some form also in the major spliceosomal assembly pathway. Spliceosomal components exhibit certain dependence on one another in the course of splicing. There is a whole range of possible spliceosomal intermediates detected, from the controversial yeast penta-snRNP (Stevens et al., 2002) across di-snRNP (Xu et al., 2004) to individual snRNPs. One can argue that there is some kind of context dependent pathways leading to assembly of the differential complexes. It might be influenced by both chromatin environment and sequence of particular pre-mRNAs. The nuclear proteins are considered to move stochastically and with a high mobility *in vivo* (Phair and Misteli, 2000) which probably enables various complexes formation and existence of both transient and stable interactions among these proteins. This is the key condition for different spliceosomal complexes formation.

7. References

Abovich, N., and Rosbash, M. (1997). Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. Cell *89*, 403-412.

Anderson, K., and Moore, M.J. (1997). Bimolecular exon ligation by the human spliceosome. Science *276*, 1712-1716.

Azubel, M., Habib, N., Sperling, R., and Sperling, J. (2006). Native spliceosomes assemble with pre-mRNA to form supraspliceosomes. J Mol Biol *356*, 955-966.

Baillat, D., Hakimi, M.A., Naar, A.M., Shilatifard, A., Cooch, N., and Shiekhattar, R. (2005). Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. Cell *123*, 265-276.

Bartels, C., Klatt, C., Luhrmann, R., and Fabrizio, P. (2002). The ribosomal translocase homologue Snu114p is involved in unwinding U4/U6 RNA during activation of the spliceosome. EMBO Rep *3*, 875-880.

Battle, D.J., Kasim, M., Yong, J., Lotti, F., Lau, C.K., Mouaikel, J., Zhang, Z., Han, K., Wan, L., and Dreyfuss, G. (2006). The SMN complex: an assembly machine for RNPs. Cold Spring Harb Symp Quant Biol *71*, 313-320.

Bedford, M.T., Reed, R., and Leder, P. (1998). WW domain-mediated interactions reveal a spliceosome-associated protein that binds a third class of proline-rich motif: the proline glycine and methionine-rich motif. Proc Natl Acad Sci U S A *95*, 10602-10607.

Behzadnia, N., Golas, M.M., Hartmuth, K., Sander, B., Kastner, B., Deckert, J., Dube, P., Will, C.L., Urlaub, H., Stark, H., *et al.* (2007). Composition and three-dimensional EM structure of double affinity-purified, human prespliceosomal A complexes. EMBO J *26*, 1737-1748.

Behzadnia, N., Hartmuth, K., Will, C.L., and Luhrmann, R. (2006). Functional spliceosomal A complexes can be assembled in vitro in the absence of a penta-snRNP. RNA *12*, 1738-1746.

Berglund, J.A., Abovich, N., and Rosbash, M. (1998). A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. Genes Dev *12*, 858-867.

Bessonov, S., Anokhina, M., Krasauskas, A., Golas, M.M., Sander, B., Will, C.L., Urlaub, H., Stark, H., and Luhrmann, R. (2010). Characterization of purified human Bact spliceosomal complexes reveals compositional and morphological changes during spliceosome activation and first step catalysis. RNA *16*, 2384-2403.

Bessonov, S., Anokhina, M., Will, C.L., Urlaub, H., and Luhrmann, R. (2008). Isolation of an active step I spliceosome and composition of its RNP core. Nature *452*, 846-850.

Bindereif, A., and Green, M.R. (1987). An ordered pathway of snRNP binding during mammalian pre-mRNA splicing complex assembly. EMBO J *6*, 2415-2424.

Boukis, L.A., Liu, N., Furuyama, S., and Bruzik, J.P. (2004). Ser/Arg-rich protein-mediated communication between U1 and U2 small nuclear ribonucleoprotein particles. J Biol Chem *279*, 29647-29653.

Cote, J., Beaudoin, J., Tacke, R., and Chabot, B. (1995). The U1 small nuclear ribonucleoprotein/5' splice site interaction affects U2AF65 binding to the downstream 3' splice site. J Biol Chem *270*, 4031-4036.

Crispino, J.D., Blencowe, B.J., and Sharp, P.A. (1994). Complementation by SR proteins of pre-mRNA splicing reactions depleted of U1 snRNP. Science *265*, 1866-1869.

Darzacq, X., Jady, B.E., Verheggen, C., Kiss, A.M., Bertrand, E., and Kiss, T. (2002). Cajal body-specific small nuclear RNAs: a novel class of 2'-O-methylation and pseudouridylation guide RNAs. EMBO J *21*, 2746-2756.

Darzacq, X., Kittur, N., Roy, S., Shav-Tal, Y., Singer, R.H., and Meier, U.T. (2006). Stepwise RNP assembly at the site of H/ACA RNA transcription in human cells. J Cell Biol *173*, 207-218.

Das, R., Zhou, Z., and Reed, R. (2000). Functional association of U2 snRNP with the ATP-independent spliceosomal complex E. Mol Cell *5*, 779-787.

Deckert, J., Hartmuth, K., Boehringer, D., Behzadnia, N., Will, C.L., Kastner, B., Stark, H., Urlaub, H., and Luhrmann, R. (2006). Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions. Mol Cell Biol *26*, 5528-5543.

Donmez, G., Hartmuth, K., Kastner, B., Will, C.L., and Luhrmann, R. (2007). The 5' end of U2 snRNA is in close proximity to U1 and functional sites of the pre-mRNA in early spliceosomal complexes. Mol Cell *25*, 399-411.

Donmez, G., Hartmuth, K., and Luhrmann, R. (2004). Modified nucleotides at the 5' end of human U2 snRNA are required for spliceosomal E-complex formation. RNA *10*, 1925-1933.

Du, H., and Rosbash, M. (2001). Yeast U1 snRNP-pre-mRNA complex formation without U1snRNA-pre-mRNA base pairing. RNA *7*, 133-142.

Du, H., and Rosbash, M. (2002). The U1 snRNP protein U1C recognizes the 5' splice site in the absence of base pairing. Nature *419*, 86-90.

Eliceiri, G.L., and Sayavedra, M.S. (1976). Small RNAs in the nucleus and cytoplasm of HeLa cells. Biochem Biophys Res Commun *72*, 507-512.

Ellis, J.D., Lleres, D., Denegri, M., Lamond, A.I., and Caceres, J.F. (2008). Spatial mapping of splicing factor complexes involved in exon and intron definition. J Cell Biol *181*, 921-934.

Forch, P., Merendino, L., Martinez, C., and Valcarcel, J. (2003). U2 small nuclear ribonucleoprotein particle (snRNP) auxiliary factor of 65 kDa, U2AF65, can promote U1 snRNP recruitment to 5' splice sites. Biochem J *372*, 235-240.

Frilander, M.J., and Steitz, J.A. (1999). Initial recognition of U12-dependent introns requires both U11/5' splice-site and U12/branchpoint interactions. Genes Dev *13*, 851-863.

Fu, X.D., and Maniatis, T. (1992). The 35-kDa mammalian splicing factor SC35 mediates specific interactions between U1 and U2 small nuclear ribonucleoprotein particles at the 3' splice site. Proc Natl Acad Sci U S A *89*, 1725-1729.

Ganot, P., Jady, B.E., Bortolin, M.L., Darzacq, X., and Kiss, T. (1999). Nucleolar factors direct the 2'-O-ribose methylation and pseudouridylation of U6 spliceosomal RNA. Mol Cell Biol *19*, 6906-6917.

Golas, M.M., Sander, B., Bessonov, S., Grote, M., Wolf, E., Kastner, B., Stark, H., and Luhrmann, R. (2010). 3D cryo-EM structure of an active step I spliceosome and localization of its catalytic core. Mol Cell *40*, 927-938.

Gornemann, J., Kotovic, K.M., Hujer, K., and Neugebauer, K.M. (2005). Cotranscriptional spliceosome assembly occurs in a stepwise fashion and requires the cap binding complex. Mol Cell *19*, 53-63.

Grainger, R.J., and Beggs, J.D. (2005). Prp8 protein: at the heart of the spliceosome. RNA *11*, 533-557.

Graveley, B.R., and Maniatis, T. (1998). Arginine/serine-rich domains of SR proteins can function as activators of pre-mRNA splicing. Mol Cell *1*, 765-771.

Grote, M., Wolf, E., Will, C.L., Lemm, I., Agafonov, D.E., Schomburg, A., Fischle, W., Urlaub, H., and Luhrmann, R. (2010). Molecular architecture of the human Prp19/CDC5L complex. Mol Cell Biol *30*, 2105-2119.

Guth, S., and Valcarcel, J. (2000). Kinetic role for mammalian SF1/BBP in spliceosome assembly and function after polypyrimidine tract recognition by U2AF. J Biol Chem *275*, 38059-38066.

Hawkins, J.D. (1988). A survey on intron and exon lengths. Nucleic Acids Res *16*, 9893-9908.

Heiner, M., Hui, J., Schreiner, S., Hung, L.H., and Bindereif, A. (2010). HnRNP L-mediated regulation of mammalian alternative splicing by interference with splice site recognition. RNA Biol *7*, 56-64.

Hertel, K.J. (2008). Combinatorial control of exon recognition. J Biol Chem *283*, 1211-1215.

Hoffman, B.E., and Grabowski, P.J. (1992). U1 snRNP targets an essential splicing factor, U2AF65, to the 3' splice site by a network of interactions spanning the exon. Genes Dev *6*, 2554-2568.

Hoskins, A.A., Friedman, L.J., Gallagher, S.S., Crawford, D.J., Anderson, E.G., Wombacher, R., Ramirez, N., Cornish, V.W., Gelles, J., and Moore, M.J. (2011). Ordered and dynamic assembly of single spliceosomes. Science *331*, 1289-1295.

Huranova, M., Ivani, I., Benda, A., Poser, I., Brody, Y., Hof, M., Shav-Tal, Y., Neugebauer, K.M., and Stanek, D. (2010). The differential interaction of

snRNPs with pre-mRNA reveals splicing kinetics in living cells. J Cell Biol *191*, 75-86.

Huranova, M., Jablonski, J.A., Benda, A., Hof, M., Stanek, D., and Caputi, M. (2009). In vivo detection of RNA-binding protein interactions with cognate RNA sequences by fluorescence resonance energy transfer. RNA *15*, 2063-2071.

Champion-Arnaud, P., Gozani, O., Palandjian, L., and Reed, R. (1995). Accumulation of a novel spliceosomal complex on pre-mRNAs containing branch site mutations. Mol Cell Biol *15*, 5750-5756.

Cho, S., Hoang, A., Sinha, R., Zhong, X.Y., Fu, X.D., Krainer, A.R., and Ghosh, G. (2011). Interaction between the RNA binding domains of Ser-Arg splicing factor 1 and U1-70K snRNP protein determines early spliceosome assembly. Proc Natl Acad Sci U S A *108*, 8233-8238.

Chusainow, J., Ajuh, P.M., Trinkle-Mulcahy, L., Sleeman, J.E., Ellenberg, J., and Lamond, A.I. (2005). FRET analyses of the U2AF complex localize the U2AF35/U2AF65 interaction in vivo and reveal a novel self-interaction of U2AF35. RNA *11*, 1201-1214.

Johnson, T.L., and Abelson, J. (2001). Characterization of U4 and U6 interactions with the 5' splice site using a S. cerevisiae in vitro trans-splicing system. Genes Dev *15*, 1957-1970.

Karaduman, R., Dube, P., Stark, H., Fabrizio, P., Kastner, B., and Luhrmann, R. (2008). Structure of yeast U6 snRNPs: arrangement of Prp24p and the LSm complex as revealed by electron microscopy. RNA *14*, 2528-2537.

Kent, O.A., and MacMillan, A.M. (2002). Early organization of pre-mRNA during spliceosome assembly. Nat Struct Biol *9*, 576-581.

Kent, O.A., Reayi, A., Foong, L., Chilibeck, K.A., and MacMillan, A.M. (2003). Structuring of the 3' splice site by U2AF65. J Biol Chem *278*, 50572-50577.

Kent, O.A., Ritchie, D.B., and Macmillan, A.M. (2005). Characterization of a U2AF-independent commitment complex (E') in the mammalian spliceosome assembly pathway. Mol Cell Biol *25*, 233-240.

Kiss, A.M., Jady, B.E., Darzacq, X., Verheggen, C., Bertrand, E., and Kiss, T. (2002). A Cajal body-specific pseudouridylation guide RNA is composed of two box H/ACA snoRNA-like domains. Nucleic Acids Res *30*, 4643-4649.

Konarska, M.M., Vilardell, J., and Query, C.C. (2006). Repositioning of the reaction intermediate within the catalytic center of the spliceosome. Mol Cell *21*, 543-553.

Konforti, B.B., and Konarska, M.M. (1994). U4/U5/U6 snRNP recognizes the 5' splice site in the absence of U2 snRNP. Genes Dev *8*, 1962-1973.

Kotlajich, M.V., Crabb, T.L., and Hertel, K.J. (2009). Spliceosome assembly pathways for different types of alternative splicing converge during commitment to splice site pairing in the A complex. Mol Cell Biol *29*, 1072-1082.

Kramer, A., Gruter, P., Groning, K., and Kastner, B. (1999). Combined biochemical and electron microscopic analyses reveal the architecture of the mammalian U2 snRNP. J Cell Biol *145*, 1355-1368.

Kuhn, A.N., Reichl, E.M., and Brow, D.A. (2002). Distinct domains of splicing factor Prp8 mediate different aspects of spliceosome activation. Proc Natl Acad Sci U S A *99*, 9145-9149.

Lacadie, S.A., and Rosbash, M. (2005). Cotranscriptional spliceosome assembly dynamics and the role of U1 snRNA:5'ss base pairing in yeast. Mol Cell *19*, 65-75.

Lange, T.S., and Gerbi, S.A. (2000). Transient nucleolar localization Of U6 small nuclear RNA in Xenopus Laevis oocytes. Mol Biol Cell *11*, 2419-2428.

Lardelli, R.M., Thompson, J.X., Yates, J.R., 3rd, and Stevens, S.W. (2010). Release of SF3 from the intron branchpoint activates the first step of pre-mRNA splicing. RNA *16*, 516-528.

Li, Y., and Blencowe, B.J. (1999). Distinct factor requirements for exonic splicing enhancer function and binding of U2AF to the polypyrimidine tract. J Biol Chem *274*, 35074-35079.

Luhrmann, R., and Stark, H. (2009). Structural mapping of spliceosomes by electron microscopy. Curr Opin Struct Biol *19*, 96-102.

Makarova, O.V., Makarov, E.M., Urlaub, H., Will, C.L., Gentzel, M., Wilm, M., and Luhrmann, R. (2004). A subset of human 35S U5 proteins, including Prp19, function prior to catalytic step 1 of splicing. EMBO J *23*, 2381-2391.

Malca, H., Shomron, N., and Ast, G. (2003). The U1 snRNP base pairs with the 5' splice site within a penta-snRNP complex. Mol Cell Biol *23*, 3442-3455.

Maroney, P.A., Romfo, C.M., and Nilsen, T.W. (2000). Functional recognition of 5' splice site by U4/U6.U5 tri-snRNP defines a novel ATP-dependent step in early spliceosome assembly. Mol Cell *6*, 317-328.

Matlin, A.J., Clark, F., and Smith, C.W. (2005). Understanding alternative splicing: towards a cellular code. Nat Rev Mol Cell Biol *6*, 386-398.

Michaud, S., and Reed, R. (1991). An ATP-independent complex commits pre-mRNA to the mammalian spliceosome assembly pathway. Genes Dev *5*, 2534-2546.

Michaud, S., and Reed, R. (1993). A functional association between the 5' and 3' splice site is established in the earliest prespliceosome complex (E) in mammals. Genes Dev *7*, 1008-1020.

Narayanan, U., Achsel, T., Luhrmann, R., and Matera, A.G. (2004). Coupled in vitro import of U snRNPs and SMN, the spinal muscular atrophy protein. Mol Cell *16*, 223-234.

Nottrott, S., Urlaub, H., and Luhrmann, R. (2002). Hierarchical, clustered protein interactions with U4/U6 snRNA: a biochemical role for U4/U6 proteins. EMBO J *21*, 5527-5538.

Pandya-Jones, A., and Black, D.L. (2009). Co-transcriptional splicing of constitutive and alternative exons. RNA *15*, 1896-1908.

Patel, S.B., and Bellini, M. (2008). The assembly of a spliceosomal small nuclear ribonucleoprotein particle. Nucleic Acids Res *36*, 6482-6493.

Phair, R.D., and Misteli, T. (2000). High mobility of proteins in the mammalian cell nucleus. Nature *404*, 604-609.

Pomeranz Krummel, D.A., Oubridge, C., Leung, A.K., Li, J., and Nagai, K. (2009). Crystal structure of human spliceosomal U1 snRNP at 5.5 A resolution. Nature *458*, 475-480.

Price, S.R., Evans, P.R., and Nagai, K. (1998). Crystal structure of the spliceosomal U2B"-U2A' protein complex bound to a fragment of U2 small nuclear RNA. Nature *394*, 645-650.

Query, C.C., Moore, M.J., and Sharp, P.A. (1994). Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model. Genes Dev *8*, 587-597.

Reed, R. (1990). Protein composition of mammalian spliceosomes assembled in vitro. Proc Natl Acad Sci U S A *87*, 8031-8035.

Reich, C.I., VanHoy, R.W., Porter, G.L., and Wise, J.A. (1992). Mutations at the 3' splice site can be suppressed by compensatory base changes in U1 snRNA in fission yeast. Cell *69*, 1159-1169.

Rino, J., Desterro, J.M., Pacheco, T.R., Gadella, T.W., Jr., and Carmo-Fonseca, M. (2008). Splicing factors SF1 and U2AF associate in extraspliceosomal complexes. Mol Cell Biol *28*, 3045-3057.

Rollenhagen, C., and Pante, N. (2006). Nuclear import of spliceosomal snRNPs. Can J Physiol Pharmacol *84*, 367-376.

Romfo, C.M., and Wise, J.A. (1997). Both the polypyrimidine tract and the 3' splice site function prior to the first step of splicing in fission yeast. Nucleic Acids Res *25*, 4658-4665.

Roybal, G.A., and Jurica, M.S. (2010). Spliceostatin A inhibits spliceosome assembly subsequent to prespliceosome formation. Nucleic Acids Res *38*, 6664-6672.

Sander, B., Golas, M.M., Makarov, E.M., Brahms, H., Kastner, B., Luhrmann, R., and Stark, H. (2006). Organization of core spliceosomal components U5 snRNA loop I and U4/U6 Di-snRNP within U4/U6.U5 Tri-snRNP as revealed by electron cryomicroscopy. Mol Cell *24*, 267-278.

Segref, A., Mattaj, I.W., and Ohno, M. (2001). The evolutionarily conserved region of the U snRNA export mediator PHAX is a novel RNA-binding domain that is essential for U snRNA export. RNA *7*, 351-360.

Sharma, S., Falick, A.M., and Black, D.L. (2005). Polypyrimidine tract binding protein blocks the 5' splice site-dependent assembly of U2AF and the prespliceosomal E complex. Mol Cell *19*, 485-496.

Sharma, S., Maris, C., Allain, F.H., and Black, D.L. (2011). U1 snRNA directly interacts with polypyrimidine tract-binding protein during splicing repression. Mol Cell *41*, 579-588.

Shen, H., Kan, J.L., and Green, M.R. (2004). Arginine-serine-rich domains bound at splicing enhancers contact the branchpoint to promote prespliceosome assembly. Mol Cell *13*, 367-376.

Schaal, T.D., and Maniatis, T. (1999). Multiple distinct splicing enhancers in the protein-coding sequences of a constitutively spliced pre-mRNA. Mol Cell Biol *19*, 261-273.

Schneider, M., Will, C.L., Anokhina, M., Tazi, J., Urlaub, H., and Luhrmann, R. (2010). Exon definition complexes contain the tri-snRNP and can be directly converted into B-like precatalytic splicing complexes. Mol Cell *38*, 223-235.

Spadaccini, R., Reidt, U., Dybkov, O., Will, C., Frank, R., Stier, G., Corsini, L., Wahl, M.C., Luhrmann, R., and Sattler, M. (2006). Biochemical and NMR analyses of an SF3b155-p14-U2AF-RNA interaction network involved in branch point definition during pre-mRNA splicing. RNA *12*, 410-425.

Spector, D.L., and Lamond, A.I. (2011). Nuclear speckles. Cold Spring Harb Perspect Biol *3*.

Sperling, J., Azubel, M., and Sperling, R. (2008). Structure and function of the Pre-mRNA splicing machine. Structure *16*, 1605-1615.

Spiluttini, B., Gu, B., Belagal, P., Smirnova, A.S., Nguyen, V.T., Hebert, C., Schmidt, U., Bertrand, E., Darzacq, X., and Bensaude, O. (2010). Splicing-independent recruitment of U1 snRNP to a transcription unit in living cells. J Cell Sci *123*, 2085-2093.

Staknis, D., and Reed, R. (1994). SR proteins promote the first specific recognition of Pre-mRNA and are present together with the U1 small nuclear ribonucleoprotein particle in a general splicing enhancer complex. Mol Cell Biol *14*, 7670-7682.

Staley, J.P., and Guthrie, C. (1999). An RNA switch at the 5' splice site requires ATP and the DEAD box protein Prp28p. Mol Cell *3*, 55-64.

Stanek, D., and Neugebauer, K.M. (2006). The Cajal body: a meeting place for spliceosomal snRNPs in the nuclear maze. Chromosoma *115*, 343-354.

Stanek, D., Pridalova-Hnilicova, J., Novotny, I., Huranova, M., Blazikova, M., Wen, X., Sapra, A.K., and Neugebauer, K.M. (2008). Spliceosomal small nuclear ribonucleoprotein particles repeatedly cycle through Cajal bodies. Mol Biol Cell *19*, 2534-2543.

Stark, H., Dube, P., Luhrmann, R., and Kastner, B. (2001). Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle. Nature *409*, 539-542.

Stark, J.M., Bazett-Jones, D.P., Herfort, M., and Roth, M.B. (1998). SR proteins are sufficient for exon bridging across an intron. Proc Natl Acad Sci U S A *95*, 2163-2168.

Sterner, D.A., Carlo, T., and Berget, S.M. (1996). Architectural limits on split genes. Proc Natl Acad Sci U S A *93*, 15081-15085.

Stevens, S.W., Ryan, D.E., Ge, H.Y., Moore, R.E., Young, M.K., Lee, T.D., and Abelson, J. (2002). Composition and functional characterization of the yeast spliceosomal penta-snRNP. Mol Cell *9*, 31-44.

Tardiff, D.F., and Rosbash, M. (2006). Arrested yeast splicing complexes indicate stepwise snRNP recruitment during in vivo spliceosome assembly. RNA *12*, 968-979.

Valcarcel, J., Gaur, R.K., Singh, R., and Green, M.R. (1996). Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA [corrected]. Science *273*, 1706-1709.

Veretnik, S., Wills, C., Youkharibache, P., Valas, R.E., and Bourne, P.E. (2009). Sm/Lsm genes provide a glimpse into the early evolution of the spliceosome. PLoS Comput Biol *5*, e1000315.

Wahl, M.C., Will, C.L., and Luhrmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. Cell *136*, 701-718.

Will, C.L., and Luhrmann, R. (2011). Spliceosome structure and function. Cold Spring Harb Perspect Biol *3*.

Wolin, S.L., and Cedervall, T. (2002). The La protein. Annu Rev Biochem *71*, 375-403.

Wu, J.Y., and Maniatis, T. (1993). Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. Cell *75*, 1061-1070.

Wu, S., Romfo, C.M., Nilsen, T.W., and Green, M.R. (1999). Functional recognition of the 3' splice site AG by the splicing factor U2AF35. Nature *402*, 832-835.

Xu, Y.Z., Newnham, C.M., Kameoka, S., Huang, T., Konarska, M.M., and Query, C.C. (2004). Prp5 bridges U1 and U2 snRNPs and enables stable U2 snRNP association with intron RNA. EMBO J *23*, 376-385.

Zamore, P.D., Patton, J.G., and Green, M.R. (1992). Cloning and domain structure of the mammalian splicing factor U2AF. Nature *355*, 609-614.

Zhou, Z., Licklider, L.J., Gygi, S.P., and Reed, R. (2002). Comprehensive proteomic analysis of the human spliceosome. Nature *419*, 182-185.

Zuo, P., and Maniatis, T. (1996). The splicing factor U2AF35 mediates critical protein-protein interactions in constitutive and enhancer-dependent splicing. Genes Dev *10*, 1356-1368.